Extracting Spatiotemporal Word and Semantic Representations from Multiscale Neurophysiological Recordings in Humans

Dr. Sydney Cash Alexander Mark Chan

Extracting Spatiotemporal Word and Semantic Representations from Multiscale Neurophysiological Recordings in Humans

ABSTRACT

With the recent advent of neuroimaging techniques, the majority of the research studying the neural basis of language processing has focused on the localization of various lexical and semantic functions. Unfortunately, the limited time resolution of functional neuroimaging prevents a detailed analysis of the dynamics involved in word recognition, and the hemodynamic basis of these techniques prevents the study of the underlying neurophysiology. Compounding this problem, current techniques for the analysis of high-dimensional neural data are mainly sensitive to large effects in a small area, preventing a thorough study of the distributed processing involved for representing semantic knowledge. This thesis demonstrates the use of multivariate machine-learning techniques for the study of the neural representation of semantic and speech information in electro/magneto-physiological recordings with high temporal resolution. Support vector machines (SVMs) allow for the decoding of semantic category and word-specific information from non-invasive electroencephalography (EEG) and magnetoenecephalography (MEG) and demonstrate the consistent, but spatially and temporally distributed nature of such information. Moreover, the anteroventral temporal lobe (avTL) may be important for coordinating these distributed representations, as supported by the presence of supramodal category-specific information in intracranial recordings from the avTL as early as 150ms after auditory or visual word presentation. Finally, to study the inputs to this lexico-semantic system, recordings from a high density microelectrode array in anterior superior temporal gyrus (aSTG) are obtained, and the recorded spiking activity demonstrates the presence of single neurons that respond specifically to speech sounds. The successful decoding of word identity from this firing rate information suggests that the aSTG may be involved in the population coding of acousto-phonetic speech information that is likely on the pathway for mapping speech-sounds to meaning in the avTL. The feasibility of extracting semantic and phonological information from multichannel neural recordings using machine learning techniques provides a powerful method for studying language using large datasets and has potential implications for the development of fast and intuitive communication prostheses.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

A1 primary auditory cortex
ALS amyotrophic lateral sclerosis
ANOVA analysis of variance
ASR automatic speech recognition
aSTG anterior superior temporal gyrus
aTL anterior temporal lobe
avTL anteroventral temporal lobe
BOLD blood oxygentation level dependent
ColS collateral sulcus
CSD current source density
CT computed tomography
CV/CVC consonant-vowel/consonant-vowel-consonant
DI abstract-concrete word task
DTI diffusion tensor imaging
EEG electroencephalography
EKG electrocardiography
EOG electrooculography
ER/ERc entorhinal cortex
ERP event related potential
F_0, F_1, F_2 fundamental frequency, first formant, second formant
FFT fast Fourier transform
FG fusiform gyrus
fMRI functional magnetic resonance imaging
HC hippocampus
HGP high gamma power
HMM hidden markov model
iEEG intracranial electrodencephalography
IFG inferior frontal gyrus
IRB institutional review board
ISI interspike interval/interstimulus interval
IT/ITc inferotemporal cortex
ITG inferior temporal gyrus
LDA linear discriminant analysis
LFP local field potential
MEG magnetoencephalography
MFCC mel frequency ceptral coefficients
MRI magnetic resonance imaging
MTG middle temporal gyrus
MUA multi-unit activity
N400 negative evoked response at 400ms
NBC naïve Bayes classifier
OTS occipito-temporal sulcus
P300 positive evoked response at 300ms

PR/PRc perirhinal cortex
PET positron emission tomography
PN picture naming task
pSTG posterior superior temporal gyrus
PSTH peristimulus time histogram
rTMS repetitive transcranial magnetic stimulation
SQUID superconducting quantum interference device
SA auditory size judgment task
SOA stimulus onset asynchrony
STG superior temporal gyrus
STRF spectrotemporal receptive field
STS superior temporal sulcus
SUA single unit activity
SV visual size judgment task
SVM support vector machine
V1 primary visual cortex
VWFA visual word form area
WM word memory task
WN word-noise task

RELATED PUBLICATIONS

CHAN, A. M., HALGREN, E., CARLSON, C., DEVINSKY, O., DOYLE, W., KUZNIECKY, R., THESEN, T., WANG, C., SCHOMER, D. L., ESKANDAR, E. & CASH, S. S. Year. Decoding multiscale word and category-specific spatiotemporal representations from intracranial EEG. *In:* Proceedings of the Computational and System Neuroscience Conference, 2010 Salt Lake City, UT.

CHAN, A. M., HALGREN, E., MARINKOVIC, K. & CASH, S. S. 2011. Decoding word and category-specific spatiotemporal representations from MEG and EEG. *Neuroimage,* 54**,** 3028-39.

CHAN, A. M., BAKER, J. M., ESKANDAR, E., SCHOMER, D., ULBERT, I., MARINKOVIC, K., CASH, S. S. & HALGREN, E. 2011. First-Pass Selectivity for Semantic Categories in Human Anteroventral Temporal Lobe. *J Neurosci,* 31**,** 18119-29.

DYKSTRA, A. R., CHAN, A. M., QUINN, B. T., ZEPEDA, R., KELLER, C. J., CORMIER, J., MADSEN, J. R., ESKANDAR, E. N. & CASH, S. S. 2012. Individualized localization and cortical surface-based registration of intracranial electrodes. *Neuroimage,* 59**,** 3563-3570.

ACKNOWLEDGEMENTS

Although the title page of this dissertation might suggest that I performed this work single-handedly, the reality is that none of it would have been possible without the support, help, and guidance of my wonderful family, friends, and colleagues. I sincerely thank everyone I have interacted with over the last six years who has helped and guided my journey through graduate school. Many members of Dr. Sydney Cash's lab at MGH, Dr. Eric Halrgen's lab at UCSD, and Dr. Thomas Thesen's lab at NYU have directly contributed to the work I present here and provided critical feedback. In particular, Andy Dykstra, Justine Cormier, Rodrigo Zepeda, Corey Keller, Jake Donoghue, Nima Dehghani, Inna Sukhotinsky, Omar Ahmed, Vinay Jayaram, Jason Naftulin, Mia Borzello, and Kristof Giber have never hesitated to assist me, constantly provided me with inspiration, and have made my time at MGH rewarding and enjoyable. I would especially like to thank my advisors Sydney Cash and Eric Halgren, as well as the rest of my thesis committee, who have guided me through this entire process, challenged me intellectually, and helped me grow as a scientist and engineer. I am grateful for the patience and willingness of the many patients and subjects who participated in these studies and provided the invaluable data that is presented here. Thanks to the EEG technicians, Kristy Tripp, Kara Houghton, Melissa Murphy and Paul Dionne, who spent countless hours helping me set up recordings in patient rooms at both MGH and BWH, and to the other clinicians, surgeons, nurses, and fellows who allowed me to observe various clinical procedures and gave me a taste of clinical neurology. In addition, my wonderful friends and classmates have provided me with countless hours of counseling, entertainment, and opportunity to unwind in between work, for which I am truly thankful. Special thanks goes to my parents, Simon and Alice Chan, and my sister, Jessica Chan, who have devoted their lives to me, and without whose encouragement and love I would not have been able to achieve so much. Finally, I am forever indebted to Fiona McAllister, whose love and support throughout this entire process has provided me with immeasurable strength, and whose ability to restore my drive and determination when things looked bleak, I will always be grateful for.

I. INTRODUCTION

Language is a uniquely human characteristic and is the fundamental method by which we are able to convey complex ideas effortlessly and efficiently. The way in which we think about and interact with the world is clearly reflected in the words we use, the meanings we ascribe to them, and the syntactic and grammatical rules we follow. Perhaps more surprisingly, the converse is also true; language itself shapes the way we think and perceive the world (Winawer et al., 2007, Gilbert et al., 2006). Understanding how the human brain generates, comprehends, and manipulates language is a crucial element in the quest for understanding and mapping human cognition.

The production and comprehension of language are highly complex processes and require many neural systems that may not traditionally be viewed as language-specific. There is often no clear boundary between the language processing system and other sensory and motor systems. For example, where visual object recognition ends and semantic representations begin is difficult to define. Similarly, understanding spoken language necessarily depends on the normal functioning of the auditory system, but where do these two pathways separate? Studying the neural basis of language is difficult because it is inextricably linked to other sensory and cognitive processing streams, and disentangling them is nearly impossible. Therefore, it is important that we do not view language as a distinct processing pathway, but instead as a system that functions on top of, and in conjunction with, all the other sensory and motor components of the brain.

Over the past several decades, the study of the neural basis of language has focused on the neuroanatomical localization of specific components of language processing (Chao et al., 1999, Caramazza, 1988, Patterson et al., 2007). These components, from auditory and visual

word recognition to construction of syntactically correct sentences, have previously been assumed to be modular in terms of function, and lie in well-circumscribed regions of cortex. This way of thinking has led to a plethora of studies examining the language deficits of lesion patients and localization of language function via neuroimaging such as functional MRI (fMRI) and positron emission tomography (PET) (Chao et al., 1999, Mummery et al., 1998, Devlin et al., 2002). The assumption of a limited number of major loci of language function has led to analysis techniques that look for the most robust responses in small regions of interest (Friston et al., 1994). Recently, however, the idea that high-order cognitive functions occur in spatially restricted areas has come into question (Haxby et al., 2001a, Ishai et al., 1999, Patterson et al., 2007). Taking a distributed view of cognitive processing, especially for representing complex information such as semantics, reduces the assumptions about spatial localization and can provide insights into the coordination of various brain areas for performing a particular task.

While localization has been a large focus of language research over the last several years, neglecting the temporal aspects of language processing limits our understanding of how the brain comprehends and produces words and sentences. Temporal information provides insights into serial, parallel, or feedback processing streams and allows for examination of measures of synchrony and causality. Consequently, neuroimaging techniques like fMRI, with their low temporal resolution, are poor tools for examining temporal relationships. By looking at a of variety electrophysiological recording techniques, such as electro- and magneto-encephalography (EEG/MEG), we can study language processing at the millisecond-level, while also examining these pathways at a range of spatial scales, from microns to centimeters.

The need to examine distributed models of language processing using a large number of recording electrodes at high temporal resolutions produces enormous quantities of data.

Traditional analysis techniques for examining neural data, such as statistical parametric mapping for fMRI (Friston et al., 1994), are typically optimized for examining large, robust differences that occur in well-defined regions. As a consequence, these techniques can easily miss subtle, but widespread activity. Multivariate techniques that are able to utilize information from a large number of sources and robustly handle high-dimensional data would provide a method by which such distributed time-series data can be examined. In particular, a number of machine learning algorithms allow for the examination of a large number of features simultaneously and can extract information that is highly distributed, both spatially and temporally. Furthermore, many of these techniques provide elegant ways of avoiding overfitting, a common problem when looking at small numbers of trials of very high dimensional data.

Additionally, if language information can be robustly extracted from electrophysiological recordings, these algorithms could provide a foundation for the development of rapid and intuitive communication prostheses. Several million people are unable to communicate adequately due to deficits at various stages of language or speech processing (Lapointe, 2005). Stroke, aphasia, and amyotrophic lateral sclerosis (ALS) are three of many diseases that may impair the ability to adequately produce speech. Understanding whether language information may be directly extracted from electrophysiological recordings may lead to language prosthetic devices that could aid these patients. Current communication prostheses based on neural signals are often slow and unintuitive, and none of them examine language signals directly. The feasibility of directly decoding language information from electrophysiology is relatively unexplored, but doing so could open important avenues of research that may aid in the development of novel communication-assistive devices.

In this dissertation, I begin by exploring the traditional models of language processing in the human brain and the classical evidence supporting such models. I discuss the current theories regarding the visual and auditory streams for processing word information, and ideas regarding the organization and representation of semantic information in the brain. I then review the major electro/magneto-physiological recording methodologies for exploring neural processing at high temporal resolutions and the machine learning techniques that are able to robustly handle such high-dimensional data.

Chapter II explores the distributed representation of semantic information through non-invasive EEG and MEG recordings. I demonstrate the feasibility of directly decoding semantic information from electro/magneto-physiological recordings using support vector machines and show that this information is consistent between modalities and individuals. I also show that these algorithms allow for the study both spatial and temporal aspects of semantic category or individual word representations, and provide a way to explore the organization of conceptual knowledge. Finally, an extensible decoding framework is proposed that may allow for the decoding of larger vocabularies in the future.

While extracranial recording techniques provide excellent coverage of the cortical surface, they lack spatial resolution, and in Chapter III, intracranial EEG (iEEG) is utilized to study semantic category-specific representations in the anteroventral temporal lobe of patients with epilepsy. By utilizing the fine spatial and temporal resolution of iEEG and microelectrodes, the temporal aspects of extracting semantic information from words are explored. Microelectrodes allow for further study of the neural correlates of category-specific representations within medial temporal structures. I also demonstrates that the use of

machine learning techniques for decoding semantic information from microelectrodes allows for quantification of information available at various times within the word processing stream.

In Chapter IV, I examine even finer spatial scales by studying the firing of single neurons that are tuned to speech-specific auditory stimuli. I study the hierarchical processing of speech stimuli at the single-unit level and explore whether the phoneme, as a construct of linguistics, is utilized by the human brain as a building block of auditory word stimuli. I also explore the link between the representation of spoken words and written words within these single unit recordings. Finally, classification algorithms are used to test whether a small population of cells carries sufficiently diverse phonetic information for the decoding of spoken words.

In the final chapter, I synthesize the results of the work presented in this dissertation. A general model of word processing, combining the data shown here with current theories and the results of other studies, is presented. I comment on potential applications of the decoding of language information and explore methods that may improve accuracies to make brain-based communication devices feasible. Finally, I will discuss important future directions for language research using multivariate machine learning techniques. Some of the data presented in this dissertation have previously been published in Chan et al. (2010), Chan et al. (2011a), Chan et al. (2011b).

1. TRADITIONAL MODELS OF LANGUAGE PROCESSING

While scholars have studied the construction, use, and origin of language for hundreds of years, it wasn't until relatively recently that we have been able to examine the underlying neural basis for language processing. Here we describe general models of language processing, ideas regarding hierarchies and streams of information, the processing of language-related

visual and auditory stimuli, and the current ideas regarding the representation of semantic information.

1.1. BROCA AND WERNICKE'S AREAS

In 1861, Paul Broca, a French surgeon, described two patients with lesions in the inferior aspect of the frontal lobe who were unable to speak (Broca, 1861). This particular type of neural impairment of language, now called Broca's aphasia or non-fluent aphasia, typically presents with non-fluent, agrammatic speech in which function words are often omitted. In general, word comprehension is usually intact. Broca claimed that this particular region in inferior frontal gyrus of the left hemisphere, now called Broca's area, was important for the production of language.

In 1874, Carl Wernicke, a German physician and anatomist, described patients with lesions near the left temporal-parietal-occipital junction who had particular difficulty in comprehending language, but maintained relatively fluent language production (Wernicke, 1874). These patients with Wernicke's, or receptive, aphasia typically have syntactically correct, fluent speech, however words are often misused or invented, and sentences rarely make sense semantically. Furthermore, these patients are typically unable to comprehend language in either spoken or written form.

These early lesion studies suggested that inferior perisylvian frontal areas may be important for language production while posterior temporal-parietal areas could be important for language comprehension. This became and remained the prevalent view of language processing for many years (Grodzinsky et al., 1999, Caramazza, 1988). Much later, theories emerged that suggested Broca's region was instead crucial for syntactical processing while

Wernicke's region was important in semantic representations (Grodzinsky, 1986, Caramazza, 1988, Caplan, 2006). This hypothesis would potentially explain the aggramatic speech of Broca's aphasics whilst also explaining the poor comprehension and "word salad" effect observed in Wernicke's aphasics. It is now fairly apparent that both of these regions carry out highly complex processing tasks, and it has remained difficult to fully characterize their exact functions (Sahin et al., 2009).

The development of functional neuroimaging, such as PET and fMRI, has allowed for more precise localization of language-related activity in healthy subjects. However, the localization of language functions based on lesion studies and neuroimaging can only lead to as much as, what Alfonso Caramazza termed, a "modern phrenology" (Caramazza, 1988), in which very specific functions are attributed to highly circumscribed brain areas using techniques slightly more modern than skull measurements. To ignore the possibility of distributed processing or neglect the temporal aspects of language processing limits our ability to understand how the brain processes and produces words and sentences.

1.2. LANGUAGE PROCESSING STREAMS

The idea of a processing "stream" in which information flows from one area to the next is ubiquitous in neuroscience research, and there are significant amounts of evidence that many of the basic functions the brain performs are structured in such a fashion (Romanski et al., 1999, Milner and Goodale, 2006, Hickok and Poeppel, 2007, Saur et al., 2008, Arnott et al., 2004, Halgren et al., 2006, Hickok and Poeppel, 2004, Rauschecker and Scott, 2009, Spitsyna et al., 2006, Ungerleider and Haxby, 1994). A processing stream structure has been observed in both the visual and auditory systems in which low level information (spots of light or simple frequency elements) enters into primary sensory areas and is subsequently passed and

processed by downstream areas to extract higher order information such as visual/auditory object identity (Rauschecker, 1998, Hubel and Wiesel, 1962).

Furthermore, for a given modality, numerous studies have suggested that information is processed by multiple parallel streams depending on the type of information that needs to be extracted. For example, in the visual system, a ventral "what" stream is believed to extract information regarding visual object identity, while the dorsal "where" stream is thought to extract information regarding spatial location and motion (Mishkin et al., 1983, Ungerleider and Haxby, 1994). A similar dual stream hypothesis also exists for auditory information, in which a ventral stream identifies sounds, while a dorsal stream maps sound to location (Arnott et al., 2004, Romanski et al., 1999, Tian et al., 2001).

This idea of a stream of processing has also been applied to the processing of language information. In this case, it is not the flow of a single type of sensory information through a processing pathway, but many types of information are fed into the language processing "stream." The commonality between these different inputs is that they all contain coded lexical and semantic information. One of the most common views of language processing streams is summarized in Figure I-1.

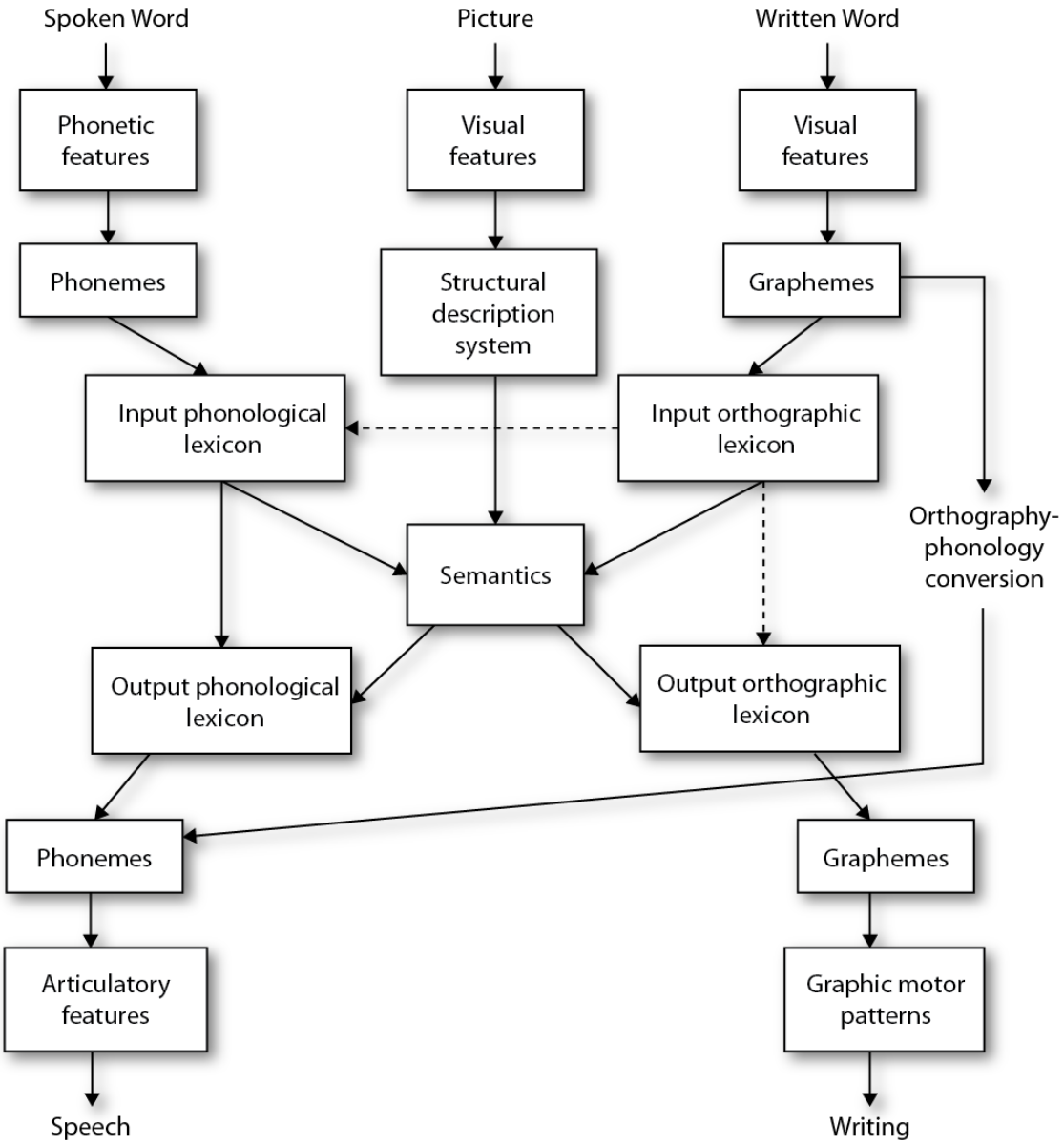

FIGURE I-1: MODEL OF LANGUAGE PROCESSING STREAMS

A common view of language processing, adapted from Martin (2003). In this model, language information flows in a hierarchical stream from which simple features are combined to form more complex objects until it leads to the activation of lexical identity and semantics. It is unclear whether the lexicons and phonological/orthographic representations are separate or common between input and output systems. Furthermore, conversions from orthography to phonology are also unclear.

In this model of single word processing, sensory input can come in the form of a spoken word, a written word, or a picture. From these inputs, either acoustic-phonetic or visual

features are extracted by early sensory processing areas from which more complex linguistic items, such as phonemes and graphemes, are constructed. Several of these objects presumably combine to form the full lexical representation of an auditory or written word. Lexical identity then provides access to the associated semantic concepts. The output pathway is similar to the input pathway, but reversed, going from a lexico-semantic concept to a phonological or orthographic lexicon by which phonemes and an articulatory plan, or graphemes and a motor plan, is generated.

While some elements of this model are likely accurate, many pieces are highly debated. For one, it is unclear whether input and output representations of lexical information (phonological and orthographic) are separate entities or whether a single lexical representation serves both purposes. Furthermore, while phonemes are ubiquitous in the study of linguistics and language, it is unclear whether the brain actually utilizes such building blocks for the representation of auditory words. Whether there exists a common lexical representation, or whether auditory and visual representations of words are separate is also unclear. Finally, it is unknown whether particular processing steps are mandatory regardless of task demands. For example, can lexical access be performed without semantic access in cases where it is not necessary (such as repetition)? Or, does the transition from orthography to lexical access require an intermediate step of phonological recoding?

While the model shown in Figure I-1 outlines the general processing modules that may be necessary for language comprehension and production, it says nothing about the neural basis of these processing steps. Understanding the neural processing of language information is crucial for placing constraints and limitations on theories of word processing. We begin by

exploring the earliest stages of language processing that involve the extraction of phonological or orthographic information from auditory and visual signals in a hierarchical fashion.

1.3. HIERARCHICAL PROCESSING OF SPEECH

While the idea of a stream describes a pathway of information flow, it does not, by itself, describe the mechanism by which information is processed. Hierarchical processing of information within these streams has been a long-standing view of sensory processing, originating in studies of the visual system (Van Essen and Maunsell, 1983, Ungerleider and Haxby, 1994). This idea suggests that at the earliest stages of processing, very simple features are extracted from low-level information, such as points or bars of light. These features are then combined to form more complex representations, such as moving bars of light or particular shapes. Proceeding down the processing stream leads to continually greater abstraction and highly complex representations of sensory information.

Studies in humans and non-human primates have shown that the auditory system, like the visual system, is organized in a hierarchical fashion (Rauschecker, 1998, Rauschecker and Scott, 2009, Tian et al., 2001, Kumar et al., 2007, Wessinger et al., 2001). At the lowest levels, neurons in primary auditory cortex typically respond to pure tones of a very specific frequency, while higher order auditory areas utilize these simple inputs to construct more complex representations of auditory inputs that are sensitive to larger frequency bandwidths or specific rates of frequency modulation (Figure I-2) (Bitterman et al., 2008, Rauschecker, 1998, Tian and Rauschecker, 1994, Tian et al., 2001, Rauschecker, 1997). In humans, the same frequency tuning of primary auditory neurons has also been observed (Bitterman et al., 2008, Howard et al., 1996), and neuroimaging has suggested that the hierarchy of general auditory information is also present (Wessinger et al., 2001).

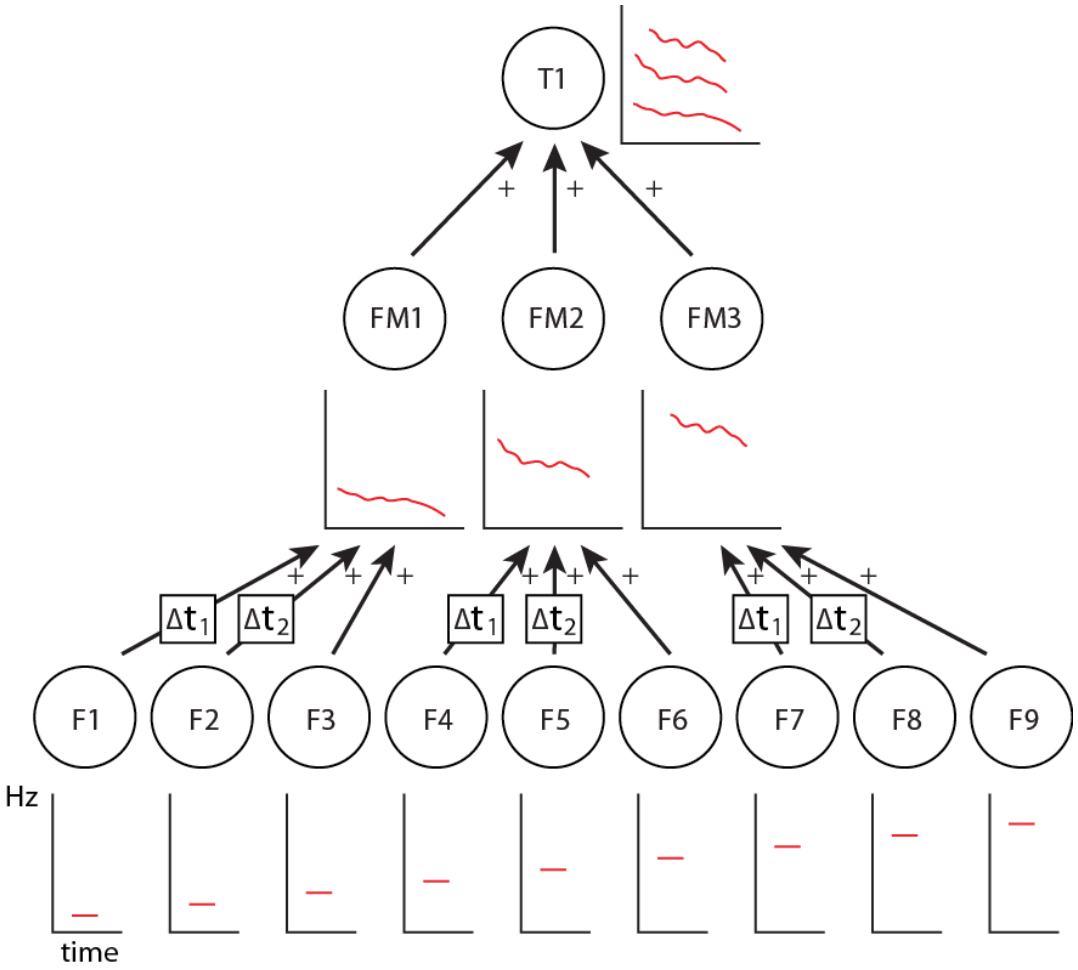

FIGURE I-2: HIERARCHICAL ORGANIZATION OF NEURONS IN THE AUDITORY SYSTEM

Illustration of the hierarchical organization of the representation of auditory signals, adapted from Rauschecker et al. (2009). Units which respond to pure frequency (F1-F9) are time shifted and summed to obtain representations of frequency modulation (FM1-FM3). These FM cells are combined to form more complex auditory representations that represent a target representation with harmonic elements (T1).

While this hierarchical structure is present for the processing of general auditory information, it has been suggested that speech information is also organized in a hierarchical fashion (Hickok and Poeppel, 2007, Rauschecker and Scott, 2009). In this case, simple auditory frequency information would be combined to form acoustic/phonetic representations, which lead to full spoken word representations (Hickok and Poeppel, 2007, Obleser et al., 2010). Unfortunately, the exact form of the representation of mid-level acoustic/phonetic information

is unknown. However, many studies suggest that individual phonemes, as studied in linguistics, are first extracted, followed by the construction of consonant-vowel (CV) pairs, or consonant-vowel-consonant (CVC) sequences (Obleser et al., 2006, Obleser et al., 2010, Jancke et al., 2002, Desai et al., 2008, Liebenthal et al., 2005). In addition, many have suggested that speech information initially enters primary auditory cortex in Heschl's Gyrus, and subsequently moves posteriorly towards Wernicke's area where phonetic information is extracted (Boatman et al., 1995, Steinschneider et al., 2011, Desai et al., 2008, Chang et al., 2010, Crone et al., 2001b, Geschwind and Levitsky, 1968, Wernicke, 1874). There is increasing evidence, however, that the anterior superior temporal lobe plays a role in phonological processing (Obleser et al., 2006, Creutzfeldt et al., 1989a, Callan et al., 2004), and Chapter IV explores the representation of speech information in single units in this area.

In further parallels to the visual system, a two stream hypothesis for speech processing has been proposed (Saur et al., 2008, Hickok and Poeppel, 2004). In this model, a dorsal stream connects superior temporal lobe to the frontal and parietal lobes via the arcuate and superior longitudinal fascicles and subserves the mapping of sound to articulation and spatial localization of speech (Hickok and Poeppel, 2004, Hickok and Poeppel, 2007). The antero-ventral stream, on the other hand, proceeds forward and inferiorly in the temporal lobe and is thought to be important for speech perception and the mapping of sound to meaning (Rauschecker and Scott, 2009, Scott et al., 2000).

Many of these previous studies have largely focused on the use of neuroimaging such as fMRI and diffusion tensor imaging (DTI) to map the neuroanatomical layout of speech processing (Saur et al., 2008). What has largely been lacking is a temporal viewpoint on the interaction of these individual processes, and discussion of feed-forward versus feedback

activation. Furthermore, while the hierarchical processing pathway of the non-human primate auditory system has been well studied down to the single unit level (Rauschecker, 1997, Rauschecker and Scott, 2009), whether the hierarchical processing of speech sounds is present at the level of single units in humans is unknown and is explored in Chapter IV.

1.4. VISUAL WORD FORM AREAS

In 1892, Dejerine described a patient with a lesion in the left inferior temporal-occipital area who had lost the ability to read or write words (Dejerine, 1892). Dejerine proposed that this area was important for carrying word-related information from the visual system to perisylvian language areas. Since then, other studies have suggested that this area is actively engaged in processing the features of visual words, rather than passively routing visual information to language areas (Binder and Mohr, 1992, Cohen et al., 2000).

This region of the posterior fusiform gyrus, that has since been named the visual word form area (VWFA), is present along the ventral visual stream at the mesial edge of the occipito-temporal sulcus. The VWFA has been shown to be consistently activated during word reading that is invariant to text size, font, or case (Dehaene et al., 2002, Dehaene et al., 2001, Warrington and Shallice, 1980, Fiez and Petersen, 1998). It has further been demonstrated that, similar to the hierarchical processing of other visual stimuli, word processing also follows a hierarchical flow (Vinckier et al., 2007, Dehaene et al., 2005). This hierarchy has been demonstrated along the length of the fusiform gyrus such that more posterior areas exhibit larger activation to false-fonts and strings of uncommon letters, while more anterior areas show greater activity to pseudowords and actual words. It is important to note that there is an ongoing debate regarding whether the VWFA is specialized for the processing of orthographic forms or whether it also processes other forms of complex visual information (Price and

Devlin, 2003, Price et al., 2003b, Cohen and Dehaene, 2004, Cohen et al., 2000, Cohen et al., 2002, Dehaene and Cohen, 2007, Dehaene et al., 2005, Dehaene et al., 2002). The idea of specialization in the language system is commented upon in Section V.1.2.

Although this visual word processing pathway is consistent in the context of the hierarchical ventral visual stream, it remains unclear how this system interacts with the other language processing networks. For example, it is unclear how a lexical item subsequently activates the semantic representation associated with that word. While it is possible that semantic concepts are directly activated from the visual-lexical representation, it has been argued that access of the phonological representation of words is an important part of visual word reading (Coltheart et al., 1993, Humphreys and Evett, 1985, Seidenberg, 1985). Some studies claim that the translation of othrographic forms to phonological information occurs before lexical access, and is a necessary step for understanding a written word (Frost, 1998, Braun et al., 2009). Others suggest a dual-route hypothesis of written word recognition such that phonological recoding of written words may occur before lexical access for infrequently encountered or novel words, while for common words, visual information leads directly to lexical access (Coltheart et al., 1993, Seidenberg, 1985). Relevant new evidence that contributes to this debate, at the single unit level, is presented in Section V.1.3.

1.5. THEORIES OF SEMANTIC KNOWLEDGE REPRESENTATION

The goal of the auditory and visual word processing streams, described previously, is to eventually feed into the lexical and semantic systems to retrieve word function and meaning. The study of the neural basis of semantic representations in humans largely began, again, by studying patients with particular brain lesions. Patients with damage to the left prefrontal cortex are often impaired when attempting to retrieve words given specific lexical or semantic

cues (Baldo and Shimamura, 1998). In addition, it was observed that some patients with temporal lobe damage have difficulty naming various objects or properties of these objects (Hodges et al., 1992, Warrington, 1975, Hart and Gordon, 1990). These studies suggested that the prefrontal and temporal cortices might be important for the storage and retrieval of lexical and semantic information.

In 1983, Warrington and McCarthy reported a patient with a cortical lesion who was impaired in producing or comprehending the names of nonliving objects (Warrington and McCarthy, 1983), and in 1984, Warrington and Shallice described four patients with herpes simplex encephalitis who demonstrated a relative impairment of knowledge of animals with a relative sparing of nonliving objects (Warrington and Shallice, 1984). These findings were subsequently observed in many other lesion patients (Damasio et al., 1996, Hillis and Caramazza, 1991, Sacchett and Humphreys, 1992, Sheridan and Humphreys, 1993, Gainotti, 1996). Most of the patients with impaired knowledge of living things had sustained damage to left temporal lobe while patients with impaired nonliving object knowledge had frontal-parietal involvement (Caramazza and Shelton, 1998, Damasio et al., 1996, Tranel et al., 1997, Mahon and Caramazza, 2009). This suggested that these two areas were important for the representation of one of these two semantic categories.

With the advent of functional neuroimaging, numerous studies have attempted to further localize the areas responsible for representation of animal or object categories. Through fMRI and PET imaging, studies have shown increased activation in lateral fusiform gyrus in response to animals (Chao et al., 1999, Martin et al., 1996, Hauk et al., 2008, Devlin et al., 2005, Perani et al., 1999, Noppeney et al., 2006, Price et al., 2003a), while objects have elicited increased activation in posterior middle temporal gyrus (Perani et al., 1999, Hauk et al.,

2008, Martin et al., 1996, Mummery et al., 1998, Mummery et al., 1996) medial fusiform gyrus (Chao et al., 1999, Devlin et al., 2005, Mechelli et al., 2006, Whatmough et al., 2002) and fronto-parietal areas (Chao and Martin, 2000). While some of the detected areas are consistent between studies, others are inconsistently found. Similar to the "modern phrenology" of localizing more general language function, it is difficult for fMRI and PET to describe the neurophysiological basis of these category representations or the order of operations for accessing semantic operations due to lack of temporal resolution and measurement of hemodynamic changes.

A number of theories have emerged that attempt to describe the organization of the semantic knowledge that might explain the double-dissociation between animal and nonliving object representations. The first theory suggests that the brain organizes conceptual knowledge into discrete evolutionarily important categories, two of which are living and nonliving objects (Caramazza and Shelton, 1998, Warrington, 1981). In other words, evolutionary pressures have led to the specialized representation of practically relevant categories such as animals and manipulable objects, and the systems that represent these different categories are distinct.

The second theory, often called the Sensory/Functional Theory, posits that the double dissociation between animal and object categories arises because the brain organizes semantic information by modality (Warrington and McCarthy, 1983, Warrington and Shallice, 1984, Gainotti, 1996). In essence, animals are largely distinguished based on their visual properties (e.g. a giraffe has four legs, a long neck, and is yellow) while manmade objects may largely be represented by their functional or associative traits (e.g. a hammer is used to hit nails). Therefore, damage to the area critical for visual representations would disproportionally

impair knowledge of animals, while damage to the motor representation might impair the knowledge of objects such as tools.

The third theory, known as the Organized Unitary Content Hypothesis, claims that the representation of a semantic concept is made up of the coactivation of all of the features and properties that define it, and that similar features are represented in nearby areas of the brain (Caramazza et al., 1990). This theory also claims that these representations do not rely on a particular sensory or motor system as in the Sensory/Functional theory. In this case, because animals tend to have many of the same features in common (e.g. have legs, eyes), and objects or tools have many of the same features in common (e.g. have handles, use with hands), damage to any specific set of feature representations would disproportionately impair knowledge of one category over another (Caramazza et al., 1990, Riddoch et al., 1988). One recent study demonstrated that by assuming the validity of this theory, the fMRI activity of various words and categories could be predicted (Mitchell et al., 2008).

None of the theories outlined above about the organization of semantic knowledge provide strong predictions about the neuroanatomical basis of this knowledge, or the processing performed when accessing semantic concepts. Most theories of the basis of semantic knowledge share the view that this information is widely distributed in the brain (Patterson et al., 2007, Martin and Chao, 2001). Semantic knowledge is closely linked to perception and action, and it is hypothesized that semantic representations may often overlap with the regions responsible for perception and action (Martin, 2007). Specifically, the ventral occipitotemporal area, within the ventral visual stream, has been implicated in the representation of object form (Chao et al., 1999, Grill-Spector et al., 2001, Haxby et al., 2001a, Shinkareva et al., 2008), the lateral posterior temporal cortex, near visual motion areas, are

thought to be important for representing objects in which motion is critical for identification (Moore and Price, 1999, Perani et al., 1995, Mummery et al., 1996, Mummery et al., 1998, Perani et al., 1999), and ventral premotor cortex has been associated with the representation of object-use-associated motor information (Chao and Martin, 2000, Grabowski et al., 1998, Grafton et al., 1997). This hypothesis is closely linked to the Sensory/Functional theory of semantic organization in which conceptual knowledge may depend on the combination of modality-specific features that are represented in the brain areas responsible for perception and action.

In addition, one recent hypothesis suggests that while semantic knowledge may be widely distributed, a semantic hub is necessary for coordinating these various areas (Patterson et al., 2007, Mummery et al., 1999, Mion et al., 2010, Lambon Ralph et al., 2010). This semantic hub would coordinate information from various areas depending on the requirements of the particular task at hand (Figure I-3). This hypothesis predicts that damage to this central processing area would lead to modality non-specific impairment of semantic knowledge. This idea has largely stemmed from the observations derived from patients with semantic dementia (Bozeat et al., 2000, Jefferies et al., 2009, Patterson et al., 2007, Lambon Ralph et al., 2007, Lambon Ralph et al., 2010, Pobric et al., 2007). These patients often show a degradation of the knowledge of people, animals, and objects that manifests as an inability to name them or make semantic generalizations. In many of these patients, degeneration of the anterior temporal lobes (aTL) is present, leading some to believe that the aTL is the location of this semantic hub. These findings have been mimicked by the use of repetitive transcranial magnetic stimulation (rTMS) to the aTL which transiently prevents the normal functioning of the targeted area (Binney et al., 2010, Lambon Ralph et al., 2009, Pobric et al., 2010a, Pobric et al., 2007, Pobric et al., 2010b).

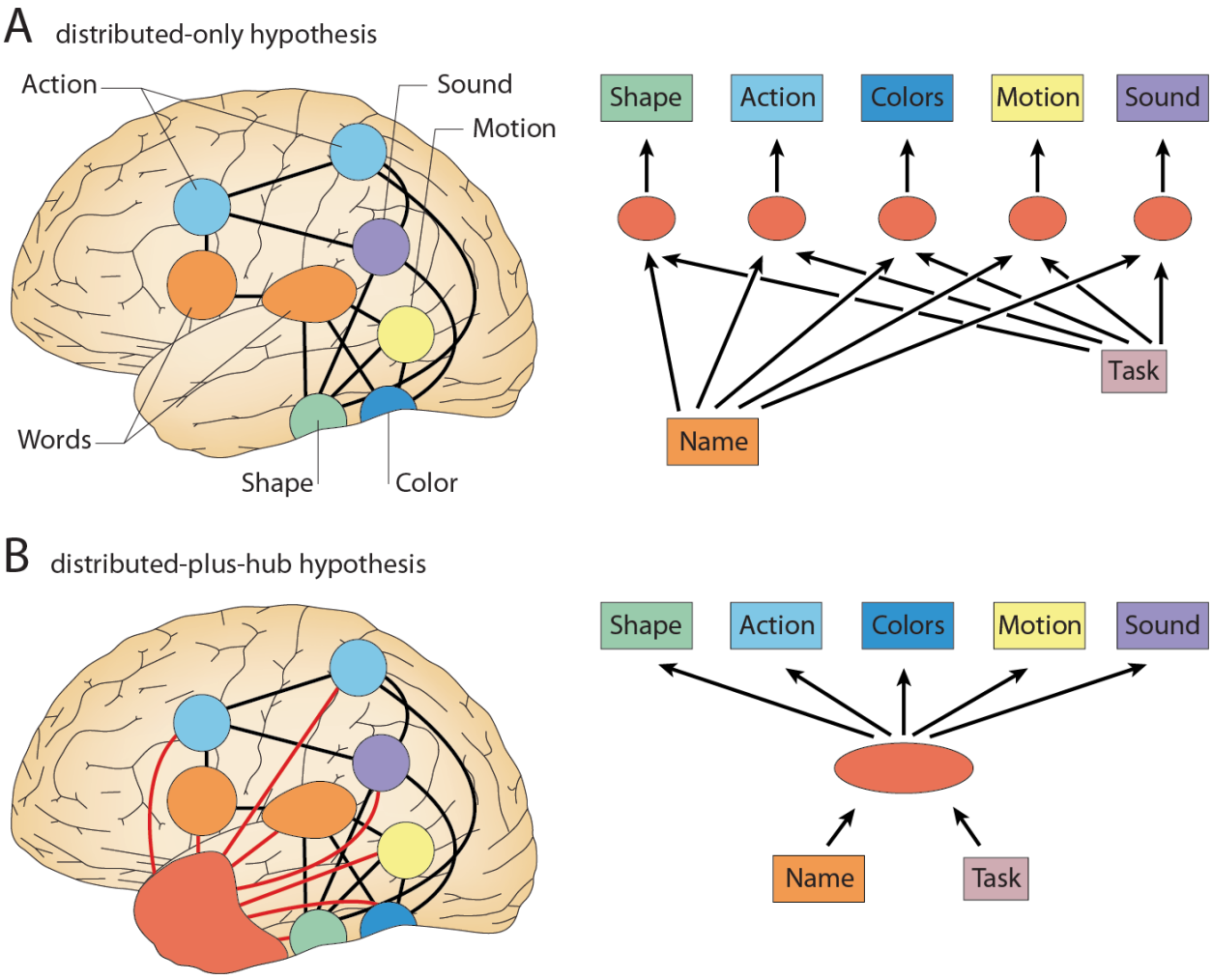

FIGURE I-3: MODEL OF SEMANTIC INFORMATION WITH OR WITHOUT HUB

Two views of semantic information representation, adapted from Patterson et al. (2007). Both views share the idea that semantic knowledge is distributed, but one hypothesis (B) suggests that a semantic hub, potentially in the anterior temporal lobe, is necessary for coordinating the different knowledge centers and allows for generalizations to be made across modality-specific information sources. Therefore, in a task that requires judgment of the size on an object when provided the spoken name of the object, the two hypotheses would predict different methods of access. In (A), the name directly activates the necessary components via the links between these areas utilizing top-down task requirements for weighting such connections. In (B), the name of the object activates the semantic hub, and combined with task requirements, the appropriate distributed features are accessed.

Most of these aforementioned studies have attempted to examine the basis of semantic knowledge by focusing on the localization of various semantic functions through lesion studies and neuroimaging (Chao et al., 1999, Chao and Martin, 2000, Perani et al., 1995, Perani et al.,

1999, Mummery et al., 1998, Mummery et al., 2000, Devlin et al., 2005). While fMRI and PET provide increased spatial resolution compared to lesions studies and allow for examination of semantic function in healthy subjects, it is difficult for either of these techniques to examine the temporal aspects of semantic representations that may allow for the study of feedforward or feedback processing. Furthermore, if semantic representations are truly distributed, utilizing univariate analysis techniques that look for large changes in a small number of variables may not be able to detect the subtle but widespread activity that we might expect. It is therefore important to examine these semantic representations using recording modalities that provide high temporal resolution, and utilize multivariate techniques to examine distributed activity.

2. RECORDING METHODOLOGIES

The advent of functional neuroimaging, in the form of fMRI and PET, has allowed for the examination of neural activity at a very high spatial resolution not previously possible with other non-invasive techniques (Chen and Ugurbil, 1999). fMRI measures the blood-oxygenation-level-dependent (BOLD) response which is a hemodynamic signal correlated to neural activity. The BOLD response, however, is not a direct measure of neural activity, and it is difficult to examine neurophysiological mechanisms of cognitive processing by only quantifying changes in blood oxygenation. Furthermore, the temporal resolution of fMRI is on the order of many seconds, making it difficult to examine the temporal aspects of neural processing, neural oscillations, or latency effects. Language processing occurs in real-time, making it important to move beyond the simple localization of language function by studying fine-scale temporal information. Additionally, magnetic susceptibility artifacts of fMRI prevents reliable imaging of activity in the anterior temporal lobes, preventing the study of semantic information and processing that may occur there.

In this dissertation, electro/magneto-physiological recording techniques are used to examine the neural mechanisms involved in language processing at multiple spatial scales while always maintaining millisecond temporal resolution. The high temporal resolution of electrophysiology confers several benefits. It allows for the analysis of timing of language information in the brain, which may allow for examination of feedforward versus feedback processing, the order of processing steps in various language pathways, and synchrony within or between neuronal populations. Furthermore, adding a temporal dimension provides a rich feature space from which machine learning algorithms may successfully extract language information. If the decoding of language signals is ever to become feasible for communication prostheses, it is important that the employed recording methodology provides sufficient temporal resolution to allow for rapid communication. In the subsequent sections, we will examine the neurophysiological basis and advantages/disadvantages of electroencephalography (EEG), magnetoencephalography (MEG), intracranial EEG (iEEG), and microelectrode recording techniques.

2.1. ELECTROENCEPHALOGRAPHY (EEG) AND MAGNETOENCEPHALOGRAPHY (MEG)

The electrical activity of the brain, recorded by electroencephalogram, was first measured by Hans Berger in 1924. Since then, it has become the predominant way in which clinicians and neuroscientists directly measure aggregate neural activity in a noninvasive fashion. EEG is typically thought to capture the synaptic activity of large populations of pyramidal cells that are arranged radially with respect to the cortical layers. Therefore, given a population of cells that is simultaneously depolarized, a voltage gradient occurs between deeper and more superficial layers that can be detected by an electrode on the scalp. EEG generally involves recording from less than ten to over 100 electrodes placed across the scalp,

but most commonly 21 electrodes organized in a standardized "10-20" layout are used (Figure I-4A). Therefore, EEG provides excellent recording coverage over most of the lateral cortical surface, however, because voltage drops with the square of the distance from the source, it is difficult to record from deeper sources (e.g. the basal ganglia, thalamus, or medial structures).

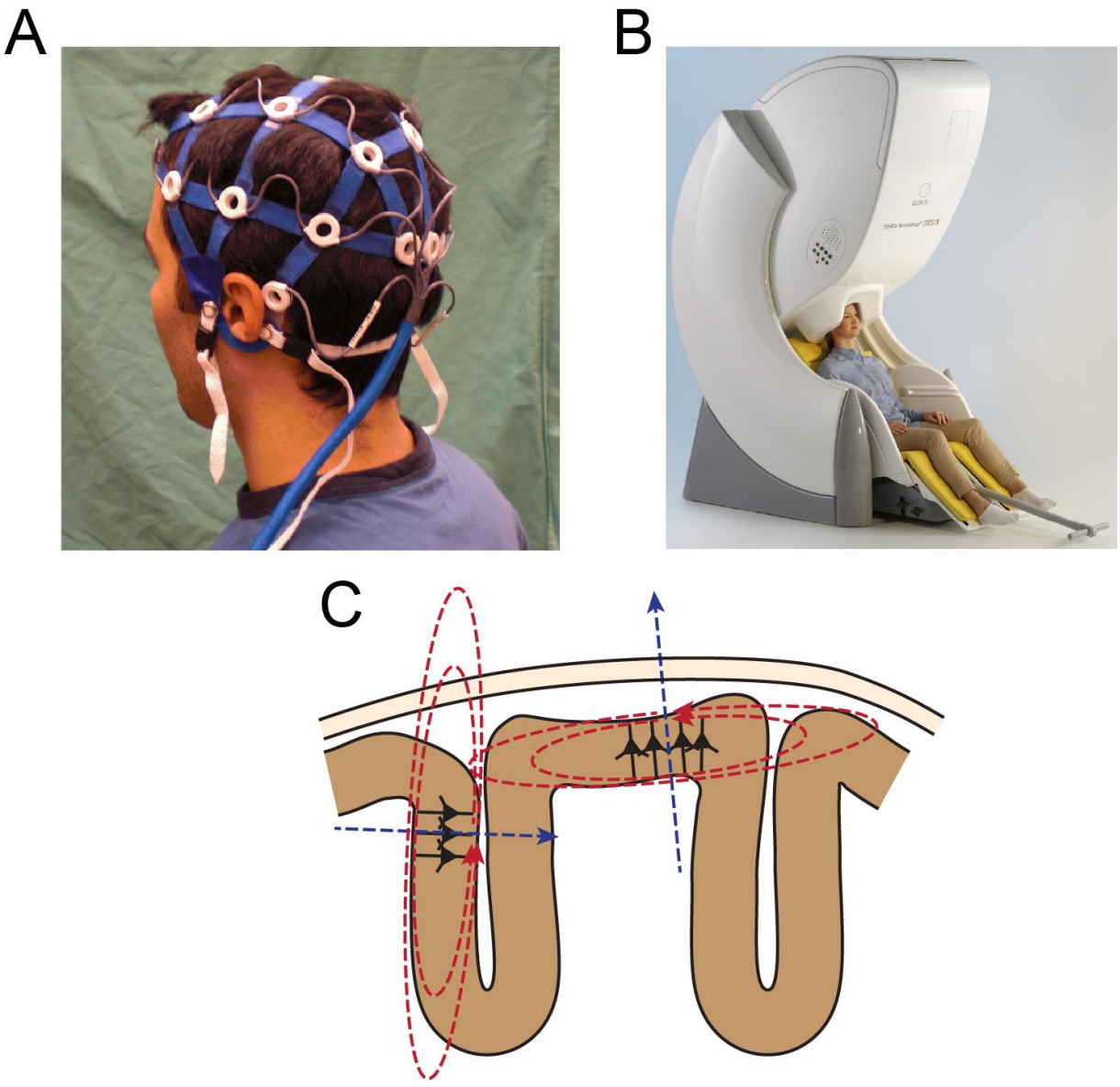

FIGURE I-4: A TYPICAL EEG AND MEG SETUP

A) A standardized EEG cap and electrodes that are typically used in cognitive experiments. A standardized 10-20 system of electrode locations usually utilizes 21 electrodes. B) A typical MEG system (Elekta, Stockholm, Sweden) with approximately 300 SQUID sensors. C) Pyramidal cells lying on a gyrus generate electrical dipoles that are oriented radially (blue arrows), allowing them

(Figure I-4 continued) to be detected outside the skull by EEG sensors, however the concurrent magnetic fields are oriented tangentially limiting their sensitivity to MEG sensors. On the other hand, cells in a sulcus have a tangentially oriented electrical dipole, but the magnetic fields propagate outside the skull allowing MEG to detect them in this case.

The lead field of EEG has been heavily debated, and there is no clear agreement on the volume of cortex which contributes to the signals recorded at a particular electrode. Most estimates place the half-sensitivity-volume (HSV) of EEG around $7cm^3$ (Malmivuo et al., 1997, Malmivuo and Suihko, 2004, Liu et al., 2002), which is far from the millimeter-scale resolution of fMRI. In addition, due to the physics of the generated electrical fields, EEG is typically more sensitive to radially oriented dipoles (e.g. cortical sources which are on the gyri), than to tangentially oriented dipoles (e.g. sources lying within the sulci) (Figure I-4C).

Since an electrical current induces a magnetic field, the same synaptic sources that produce voltage potentials also generate magnetic fields, and these fields can be recorded using magnetoencephalography (MEG) (Figure I-4B). Despite the fact that these fields are extremely small, around 10^{-12} tesla, the superconducting quantum interference device (SQUID) used in MEG is an extremely sensitive detector that is able to record such signals. MEG is typically more sensitive to tangentially oriented dipoles (e.g. activity occurring in sulci), making the information it provides complementary to the information EEG provides. The smearing effect the skull has on voltage potentials has much less influence on magnetic fields, leading to potentially better resolution of high-frequency activity and smaller lead fields. While it was traditionally believed that MEG was better able to localize cortical sources of activity than EEG using inverse models, it is unclear whether this is true if the number of sensors employed in both modalities is equal (Malmivuo et al., 1997, Malmivuo and Suihko, 2004, Cohen et al., 1990, Liu et al., 2002). High, high density EEG electrodes are rarely used, while MEG typically employs over 300 sensors.

In many cases, both magnetometers, which measure the absolute strength of the magnetic field underneath the sensor, and gradiometers, which measure changes in magnetic field in a particular direction, are used in MEG recordings. Because the MEG scanner is physically large, due to the necessary liquid helium cooling of the SQUID sensors, and a magnetically shielded room is needed, the use of MEG for communication prosthesis is unviable. Despite this, MEG provides substantial temporal resolution and coverage that makes it ideal as a first pass in studying distributed spatiotemporal semantic representations.

EEG and MEG, have excellent spatial coverage due to their extracranial nature, however, it is difficult to record from several critical brain areas using these techniques. Brain areas that have particular relevance for the representation of objects and semantic knowledge, such as the inferior and medial temporal lobe (e.g. inferotemporal cortex, entorhinal cortex, perirhinal cortex, parahippocampal gyrus, and hippocampus), cannot be reliably recorded using EEG and MEG. One way to effectively study the role of such inferior and medial areas with high temporal resolution is to record from electrodes that are placed within the skull that can access these brain regions.

2.2. INTRACRANIAL EEG (iEEG)

While EEG and MEG provide excellent temporal resolution, their spatial resolution is fairly low. Also, due to being recorded extracranially, they are susceptible to eye movement (EOG), EKG, and muscle artifacts. By using electrodes which are implanted intracranially, we are able to maintain millisecond-scale time-resolution, reliably record high frequency activity, examine much finer spatial scales of activity, and record from areas that are inaccessible by non-invasive electrophysiology (Pfurtscheller and Cooper, 1975). Similar to EEG and MEG, iEEG also records aggregate synaptic activity, however the spatial resolution of these

electrodes is much higher (approximately 1-2cm^3). Intracranial EEG (iEEG) electrodes are either subdural surface electrodes which lie on the pial surface to record from lateral cortex, or depth electrodes that penetrate the brain to record from deeper structures such as the medial temporal lobe or cingulate. Surface electrodes are typically 5mm platinum discs (2.3mm exposed diameter) embedded in a silastic sheet with a 10mm distance between contacts (Figure I-5A-B). These electrodes are often arranged in an 8x8 "grid" or 1x8 "strip" and a number of these grids or strips are often implanted to provide the necessary coverage of the cortical surface. Depth electrodes, on the other hand, consist of a long, flexible shaft with 6-8 electrodes along its length (Figure I-5C). Electrodes are platinum cylinders 1.1mm in diameter and 2.3mm long with 5mm spacing between each contact. These electrodes are typically implanted such that they penetrate the brain from the lateral surface and enter 10-15cm to record from deeper structures. In general, 4-5 depth electrodes are implanted into each hemisphere of the brain (Figure I-5D).

Due to the invasiveness of intracranial electrodes, these recordings are only performed for clinically relevant indications. In this dissertation, all recordings from iEEG are obtained from patients with medically intractable epilepsy who are undergoing presurgical evaluation for localization and eventual resection of their seizure focus. These electrodes are implanted semi-chronically (i.e. from several days to several weeks) and provide a unique opportunity to directly study neurophysiology invasively in humans. Unlike EEG and MEG, intracranial electrodes generally provide more restricted coverage of the lateral cerebral cortex, and the exact placement, number, and type of electrodes implanted is always based solely on clinical decision-making. In many cases, intracranial surface electrodes provide excellent unilateral coverage of cortex surrounding the Sylvian fissure and central sulcus (i.e. lateral temporal lobe,

anterior/inferior parietal lobe, and motor/premotor cortex). This is ideal for the study of various aspects of language that have been localized to perisylvian cortices.

Depth electrodes often allow recording from bilateral inferior temporal lobe, mesial temporal structures (e.g. hippocampus, amygdala, parahippocampal cortices), and cingulate gyrus, which are difficult to record from using extracranial electrophysiological techniques. These depth recordings (which are utilized in the analysis performed in Chapter III), are particularly valuable due to the inability of both extracranial electrophysiology as well as functional neuroimaging to record activity from these areas. In the case of fMRI, magnetic susceptibility artifacts are commonly present around the anterior and inferior temporal lobes, preventing the study of the crucial processing of semantic information that may occur in these areas.

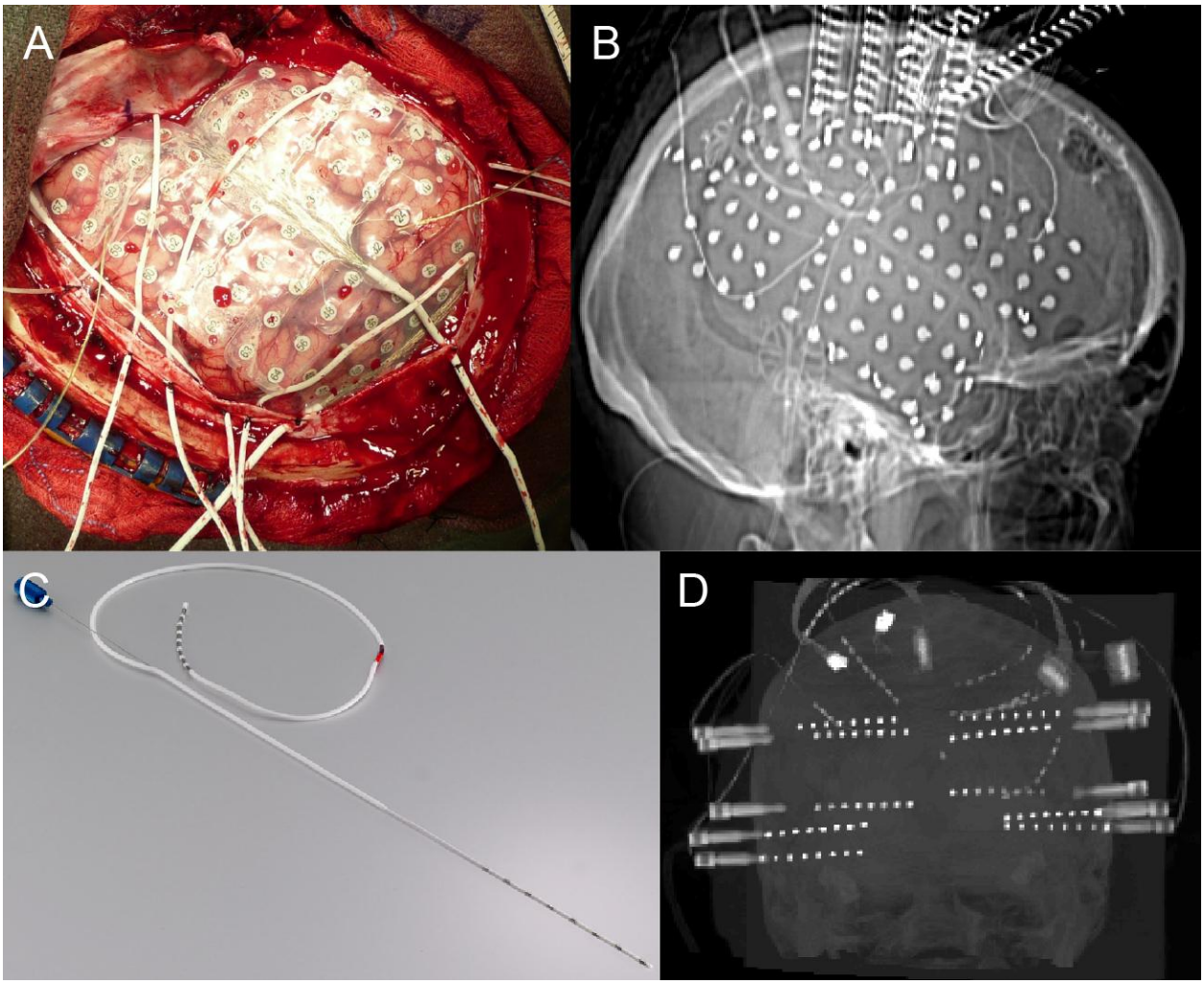

FIGURE I-5: INTRACRANIAL ELECTRODES

A) A view of the right lateral cortical surface with a 64-contact subdural grid placed over it. B) Sagittal X-ray of the implanted grid. C) An 8-contact depth electrode. D) A coronal CT image of 10 implanted 8-contact depth electrodes with 5 on each side.

2.3. MICROELECTRODES

EEG, MEG, and even iEEG all record the summed synaptic activity from large populations of cells, which can be thought of as the average input to a particular cortical area. It has been observed, however, that the average synaptic activity to a large area can be very different from the spiking activity of individual neurons (Nir et al., 2007, Rasch et al., 2008). Examining the activity of small groups of individual neurons, or even single neurons, would

allow for the characterization of single-cell responses to language stimuli and would provide a way of studying the output, rather than input, of cortical columns.

To record neural activity at such a high spatial resolution, a number of different types of penetrating microelectrode arrays have been developed (Ulbert et al., 2001, Campbell et al., 1991, Kipke et al., 2003, Kim et al., 2010). Microelectrode arrays can record local field potentials (LFPs), multiunit activity, and single unit activity from multiple contacts simultaneously, and sometimes provide interpretable spatial information. These data allow for a characterization of the neuronal inputs (synaptic activity) as well as the outputs (unit firing activity) from a small set of cells. Microelectrodes have been used extensively in animal research (Rousche and Normann, 1998, Santhanam et al., 2004, Tian and Rauschecker, 1994), but their use in humans is very limited (Truccolo et al., 2011, Hochberg et al., 2006, Fried et al., 1997). In the same population of epilepsy patients who are implanted with clinical iEEG electrodes, we have the opportunity to implant an additional microelectrode array, provided that the patient consents to the research and it is known *a priori* what region is likely to be resected. This provides a unique opportunity to study language at the single neuron level.

In this dissertation, we utilize two types of microelectrode arrays. The first is a linear microelectrode array with contacts along the shaft of a thumbtack-like probe designed by Ulbert et al. (2001) (Figure I-6A). The length of the shaft ranges from 2-4mm with 24 contacts spaced evenly along its length. This allows for the simultaneous recording of laminar activity across all cortical layers and provides a powerful way to examine incoming and outgoing information from the recorded cortex. These electrodes can also record the firing of small populations of cells, termed multi-unit activity (MUA). MUA is seen in the high frequency (>300Hz) component of these recordings, and bandpass filtering the signal followed by

rectification provides a measure of the firing of a small group of neurons. While multi-unit activity is most commonly recorded from these laminar electrodes, it is also occasionally possible to isolate the firing of single neurons, giving rise to measures of single-unit activity.

Besides measuring the firing activity in a cortical column, these laminar electrodes are also able to measure the synaptic inputs to the recorded region in the form of the current source density (CSD). The CSD is a distribution of the sources and sinks of current along the depth of the cortical layers, and can be computed as the second spatial derivative of the field potential. Current sources indicate that positive charge is flowing out of the cell at that location while current sinks indicate that positive charge is flowing into the cell (potentially depolarization). In most cases, a source and sink are paired due to the loop of current within the dendrites and extracellular media. A given sink-source pair can be generated in one of two ways: the sink can be an active sink due to an excitatory input, or the source may be an active source due to inhibitory input. The recording of simultaneous multi-unit activity can distinguish between these two possibilities. An inhibitory input will demonstrate decreases in firing, while an excitatory input will show increases in firing. We will utilize this type of electrode and analysis in Chapter III.

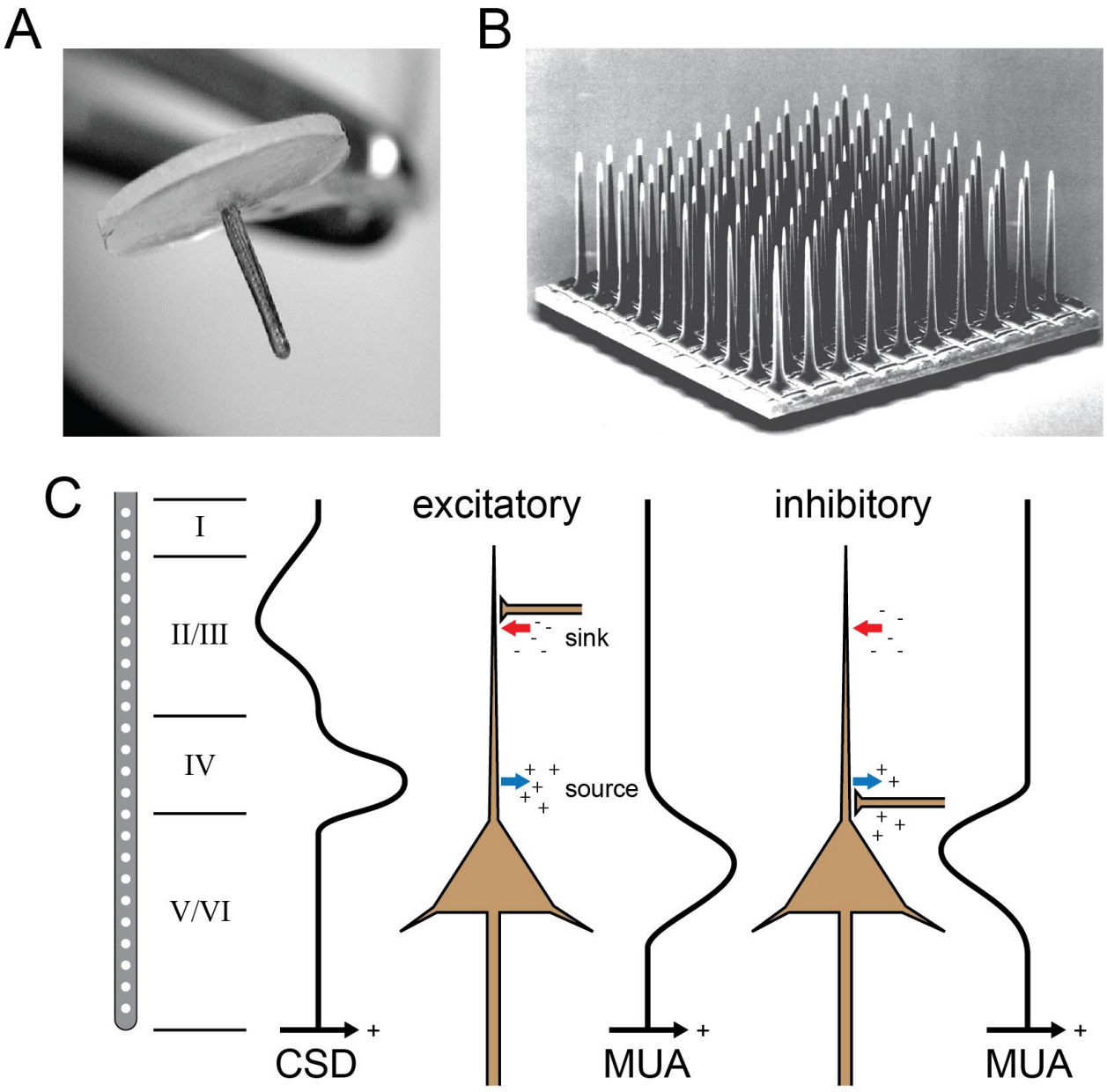

FIGURE I-6: MICROELECTRODE ARRAYS

A) The 24 contact "thumbtack" microelectrode array designed by Ulbert et al. (2001) allows for examination along the length of the cortical layers. B) The Utah Array (Blackrock Microsystems, Salt Lake City, UT) allows for the recording of LFP and single unit activity from 96 electrodes in a 4x4mm array. C) Illustration of the physiology that the laminar microelectrode captures. Given a current source density (CSD) profile, current sources and sinks can be localized within the cortical layers. In this case, a deep source occurs simultaneously with a superficial sink. This scenario can be generated by either an excitatory input superficially, or a deep inhibitory input. To distinguish between the two cases, the multi-unit activity can be recorded and would show an increase in firing near the cell body for the excitatory case, but a decrease in firing for the inhibitory case.

The second type of microelectrode is a 10x10 array of penetrating probes spaced 400μm apart, resulting in a 4mmx4mm square array (Figure I-6B). Each electrode within the array has a 20μm exposed platinum tip that allows for the recording of highly localized activity. Electrodes are 1-1.5mm in length, allowing for the recording of a large number of single neurons from layers II, III or IV. As with the laminar microelectrode, this microelectrode array (known as the Utah Array) allows for the recording of both LFP and single-unit firing, but also provides spatial information within a given cortical layer due to the regular spacing of the contacts. Chapter IV describes recordings obtained from a Utah Array in a human patient. Recordings made from both of these microelectrode arrays are sampled at 20-30kHz, providing sufficient time resolution for resolving the waveform shapes of extracellular action potentials.

3. MACHINE LEARNING TECHNIQUES

As neural recording technology and storage solutions have improved over the last several decades, it has become common to collect larger and larger neural datasets. In the case of fMRI, thousands of voxels of information are obtained per trial, while EEG and MEG obtain hundreds of channels worth of data at 100-500Hz. With the continuous 30kHz recording from the 96 channel microelectrode arrays described above, datasets easily exceed 1TB per 24hr period. To look for robust effects within such enormous datasets, a number of statistical techniques have been developed. In fMRI, statistical parametric mapping (SPM) allows for construction of a model of the activity at each voxel based on the experimental conditions (Friston et al., 1994). This allows for hypothesis testing between a number of experimental conditions, but because thousands of voxels are tested, a correction for multiple comparisons is necessary. This adjusts the final false positive rate, but also greatly reduces the sensitivity of such an analysis. This technique has been further extended to examination of the time-series

data obtained via MEG and called dynamic statistical parametric mapping (dSPM) (Dale et al., 2000), but this analysis suffers from similar drawbacks.

Both of these techniques are ubiquitous in the study of fMRI and MEG data, and because of the correction for multiple comparisons, they are best able to detect large effects in contiguous areas of cortex. These techniques are not designed to detect subtle but widespread activity that might be expected from distributed models of semantic activity. Multivariate machine learning techniques, on the other hand, are able to handle high-dimensional datasets and determine the combination of variables that provides the most discriminative power between two or more experimental conditions. These algorithms utilize statistical techniques to learn decision rules that allow for the decoding of information from a set of neural feature inputs. Besides providing a method for decoding of neural data, the generated models often allow for the quantification of information content in a set of features and can be used to explore spatiotemporal language representations from electrophysiological data. One particularly robust machine learning classifier developed over the last several years is the Support Vector Machine (SVM).

The SVM is a supervised learning algorithm that attempts to discriminate between two or more classes (e.g. different experimental conditions) based on a potentially high dimensional set of features computed for each class. Therefore, by training an SVM to discriminate between the representation of animals and objects in MEG data, we can generate a model that can later be used to predict the class membership of a novel set of neural data. While many algorithms for classification exist, support vector machines have been shown to have very good generalization performance, and have proved to work well with neural data (Shoeb et al., 2004, Mourao-Miranda et al., 2005, Shao et al., 2009, Wai Kei et al., 2008, Chan et

al., 2008). Many other classification algorithms are often based on measures of the "center" of the distributions of the given data (e.g. Fisher Discriminants). SVMs, on the other hand, belong to a class of algorithms known as "maximum margin classifiers" (Vapnik, 1995). SVMs generate a decision boundary based on the border cases that are most difficult to classify, while the majority of the points that are easy to classify are essentially ignored. This can provide very good generalization performance, helps to reduce overfitting, and makes SVMs ideal for high dimensional data sets such as the neural recordings obtained from EEG, MEG, and iEEG.

To be more formal, given a set of feature vectors and labels $(\boldsymbol{x}_i, y_i)$ where $\boldsymbol{x_i} = [x_1 \quad x_2 \quad \dots \quad x_d]^{\mathrm{T}} \in \mathbb{R}^d$ and $y_i \in \{1, -1\}$, SVMs and other learning algorithms find a mapping $f(x) \mapsto \mathbb{R}$. In the case of neural data, $\boldsymbol{x}_i$ are a set of potentially high dimensional neural data while y_i are the class labels that might correspond to "animals" and "objects". Given a new feature vector, the classifier can be used to predict the class label as $\hat{y} = \mathrm{sign}\big(f(x)\big)$. SVMs attempt to learn the function, *f*, by minimizing

$$\underset{f \in \mathcal{H}}{\operatorname{argmin}}\, C \sum_{i=1}^{n} \big(1 - y_i f(\boldsymbol{x}_i)\big)_+ + \frac{1}{2}\|f\|_{\mathcal{H}}^2$$

where, $(s)_+ = \max(s, 0)$ is the hinge loss, C is a regularization parameter, and $\mathcal{H}$ is a Reproducing Kernel Hilbert Space that limits the set of possible functions that the SVM can learn. This hinge loss penalizes all misclassified training set points, but also penalizes correctly classified points within a certain distance from the decision boundary. This loss function therefore attempts to maximize the margin by which the two classes are separated.

Because the space of functions SVMs model is a Reproducing Kernel Hilbert Space, we can view the SVM as projecting our d-dimensional data into an even higher (possibly infinite)

dimensional feature-space via a nonlinear feature map $\Phi: \mathbb{R}^d \rightarrow \mathbb{R}^m$ where $m > d$. The SVM then learns a linear boundary between the classes in this feature-space. Because training and testing an SVM only requires the computation of dot products in the feature-space, performing the transformation Φ explicitly is unnecessary. Instead, one can simply define and use the Kernel, or dot product in this high dimensional space, $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \Phi(\boldsymbol{x}_i), \Phi(\boldsymbol{x}_j) \rangle$. This linear decision boundary in the feature-space is equivalent to a non-linear decision boundary in the input space, and provides SVMs with part of their flexibility.

While SVMs often provide very good performance, the models they generate are often difficult to interpret. The Naïve Bayes Classifier (NBC) has a probabilistic formulation that is easily interpretable, and has shown surprisingly good classification performance despite its simplicity. It operates under the assumption that all of the features in the training set are independent of one another.

$$P(\boldsymbol{x}) = P(x_1, x_2, \dots, x_n) = P(x_1)P(x_2) \dots P(x_d) = \prod_{i=1}^{d} P(x_i)$$

Given a new set of features, NBCs choose the class, *C*, with the highest posterior probability. Using Bayes rule, it computes the posterior probability of each class given the set of features as

$$P(C|\boldsymbol{x}) = P(C|x_1, x_2, \dots, x_n) = \frac{P(x_1, x_2, \dots, x_d|C)P(C)}{P(x_1, x_2, \dots, x_d)} \propto P(C) \prod_{i=1}^{d} P(x_i|C)$$

and chooses the class based on

$$f(\boldsymbol{x}) = \underset{C}{\operatorname{argmax}}\, P(C) \prod_{i=1}^{d} P(x_i|C)$$

Therefore, probabilities are generated for each potential class making it relatively simple to interpret the output and intermediate steps of this algorithm. While the independence assumption is often invalid for many datasets, the Naïve Bayes Classifier often still performs well (Rish, 2001, Lewis, 1998). Unlike SVMs, however, the Naïve Bayes Classifier requires a choice of probability distribution, and this probability distribution can heavily affect the performance of the classifier. Depending on the choice of the probability distribution, Naïve Bayes Classifiers can generate nonlinear boundaries, but in the case of equal variance Gaussians, linear boundaries are generated from the product of the feature likelihoods. For high dimensional feature vectors, this may not be a problem.

While classifiers such as SVMs and NBCs do not intrinsically model time series information, we can still utilize these techniques to quantify changing amounts of information over time. Decoding features in sliding windows generates a series of accuracies that can be used to understand how much information is present at any particular point in time. In this dissertation, this technique is commonly used to determine when semantic or phonological information first appears in the recordings and features of interest.

Another machine learning technique that elegantly handles time series data is the Hidden Markov Model (HMM). HMMs are heavily used for representing phonetic information in automated speech recognition systems, and it is possible that they may also be used to decode phonological information from neural data. HMMs assume that the modeled system shifts between a set of discrete states, however these states cannot be directly observed (Figure I-7). Instead, each state emits an observable output value, or set of values, that can be represented by a probability distribution. Switching between states is determined by a set of transition probabilities that determines whether the system stays in its current state or moves

to the next state at the next time point. The Expectation-Maximization algorithm allows for the training of the parameters of an HMM given a set of data, and the forward-backward algorithm allows for the computation of the likelihood of observing a set of output values for a given model. While this dissertation mainly utilizes classification algorithms like SVMs and NBCs, I demonstrate the potential use of an HMM for decoding phonological and word-specific representations in the Conclusion, Section V.2.2.

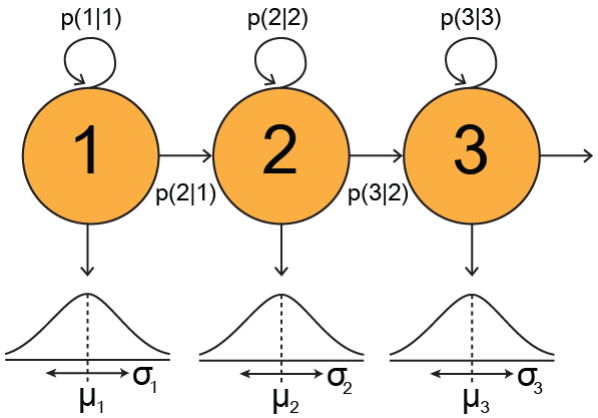

FIGURE I-7: EXAMPLE OF HIDDEN MARKOV MODEL

An example of a three state, left-to-right Hidden Markov Model. Each state S_i emits one, or multiple, output value(s) that follows a probability distribution (e.g. $\mathcal{N}(\mu_i, \sigma_i)$). The system moves between states with the indicated transition probabilities. In the case of the phonological representation of a word, each state may be one phoneme in a sequence of phonemes with the output variables being either an observed acoustic feature (when observing speech), or a set of neural information (when recording information from the brain).

Regardless of the choice of machine learning technique, all of these algorithms allow for the consideration of high dimensional data. While the classification accuracy gives a sense of the amount of information present in the underlying neural data, we can also examine the generated models to derive an understanding of which features contributed most to the classification. This provides a method by which distributed activity in the high dimensional electrophysiological data from EEG, MEG and iEEG can be quantitatively analyzed and studied.

4. COMMUNICATION PROSTHESES

Utilizing machine learning techniques to extract language information from electrophysiology not only allows for the study of distributed representations of semantic, lexical, and speech information, but may also provide the foundation by which the decoding of this information may be put to practical use. Over 1 million Americans are unable to use verbal language to communicate adequately due to disruptions in various parts of the neural pathways required for normal language production (Lapointe, 2005). Currently available communication assistance systems do not directly decode language intent. Instead, these approaches allow for the choosing of letters, words, or symbols indirectly through a secondary pathway such as motor imagery, the P300 potential, or eye movement (LaCourse and Hludik, 1990, Pfurtscheller and Neuper, 2001, Pfurtscheller et al., 1998, Farwell and Donchin, 1988). In current forms, these prostheses are slow and may be difficult to use. The P300 speller, for example, operates at speeds of 10 characters per minute or less (Donchin et al., 2000). Alternatively, one may imagine a prosthesis which could directly decode a combination of semantic, lexical, phonetic, and articulatory representations to allow entire words or concepts to be generated. This second approach is enormously challenging but may eventually provide an intuitive method for affected patients to communicate.

Direct decoding of higher order neural information for use in brain computer interfaces (BCI) has grown rapidly in the past several decades. Perhaps the most sophisticated endeavor in this regard is found in advances in motor prosthesis development (e.g. Hochberg et al., 2006, Leuthardt et al., 2006, Donoghue, 2008, Pfurtscheller and Neuper, 2001, Santhanam et al., 2006, Wilson et al., 2006, Birbaumer, 2006). Both extracranial and intracranial EEG recordings have been used in humans to provide the control signals for motor-prosthetic devices. For intracranial devices, it has been shown that the accuracy of such devices can approach 95%

when choosing between 4 targets (Santhanam et al., 2006) and a tetraplegic human patient with an implanted microelectrode array has demonstrated the ability to control a computer cursor or prosthetic hand (Hochberg et al., 2006). The ability to decode perception of visual scenes and objects from primary visual cortex via imaging techniques has also been extensively explored (Haxby et al., 2001b, Kay et al., 2008, Kamitani and Tong, 2005). These successes strongly demonstrate that essential and detailed information can be extracted from ongoing neural activity at multiple spatial and temporal scales.

Despite the success in decoding other types of neural activity, it is unknown whether machine learning techniques can extract language-specific information from electrophysiological recordings. There are several lines of research that are attempting to extract low-level information for the development of communication prostheses, but none of these methods directly examine semantic information.

Brain-computer interfaces based on the P300 evoked-potential are some of the most well-developed communication prostheses. P300 spellers are based on the "oddball" effect seen in EEG recordings and have relatively good accuracy in healthy individuals (Birbaumer, 2006, Farwell and Donchin, 1988, Kaper et al., 2004). The P300 is an event-related potential seen in EEG recordings that occurs when a rare, oddball stimulus is presented (Donchin et al., 2000). This decoding scheme allows the subject to select a letter from a grid by concentrating on the letter while each row and column of the grid is highlighted. Highlighting of the appropriate row and column evokes the P300, and this potential can be decoded if a sufficient number of repetitions are performed. Unfortunately, this type of device is relatively slow, requiring many seconds to communicate a single letter (Donchin et al., 2000).

There are several possible approaches that may allow for the development of faster and more intuitive communication prostheses. A prosthesis which decodes the articulatory motor commands used in speech is currently in the early stages of development (Brumberg et al., 2009, Guenther, 2009). This paradigm utilizes invasive intracranial recordings in order to record neural activity from speech motor areas, and attempt to decode articulatory commands to reconstruct phonetic speech information. While this approach may eventually demonstrate high levels of accuracy, it is unclear whether a prosthesis based on motor activity would be viable for patients that have impaired language ability due to higher-order functions.

While both of these techniques show promise as communication prosthesis, very little work has been done to engineer a prosthesis which is based directly on semantic language information. A prosthetic device which could decode language processing from neural activity would be highly intuitive and could possibly surpass the maximum speed of current prosthetic devices by directly decoding entire words and concepts in a single step, rather than requiring the subject to construct words from letters or sounds. If such high-order language representations are to be decoded from neural recordings, it is important that the spatiotemporal representation of semantic information be understood. We begin by exploring these distributed representations using machine learning techniques that can handle high-dimensional data.

II. DECODING DISTRIBUTED SEMANTIC REPRESENTATIONS FROM MEG AND EEG

1. INTRODUCTION

Understanding the neural basis of semantic knowledge is crucial for informing or constraining current theories regarding the organization of semantic information in the brain. Because semantic information covers an extremely broad domain of knowledge, there are many different ways for examining its representation in the brain. However, contrasting the neural differences between representing animals and manmade objects has been an extremely popular technique for studying semantics due to the robust effects observed in lesion studies (Warrington and McCarthy, 1983, Warrington and Shallice, 1984).

With the advent of functional neuroimaging techniques (e.g. PET and fMRI), numerous studies have been performed to investigate the neural basis of semantic representations. Neuroanatomical differences in the representation of specific semantic categories, especially living and non-living objects, have been seen in imaging studies (Hauk et al., 2008, Chao et al., 1999, Caramazza and Mahon, 2003, Martin and Chao, 2001, Dhond et al., 2007, Caramazza and Shelton, 1998, Shinkareva et al., 2008, Tranel et al., 1997). Despite extensive work investigating the animate/inanimate distinction, the reported results are variable from study to study (Moore and Price, 1999, Devlin et al., 2002). Most studies agree that the left posterior middle temporal gyrus is activated in response to tools and man-made objects (Perani et al., 1999, Damasio et al., 1996, Martin et al., 1996, Mummery et al., 1998, Mummery et al., 1996, Chao et al., 1999, Moore and Price, 1999), and that inferior temporal-occipital cortex is activated for animals and natural stimuli (Perani et al., 1995, Perani et al., 1999, Chao et al., 1999, Damasio et al., 1996). However, results are conflicting with regard to the medial temporal surface, left medial frontal cortex, and parietal cortex; several studies suggest

activation for animals in these areas (Damasio et al., 1996, Martin et al., 1996) while other studies find activation by man-made and non-living objects (Mummery et al., 1998, Mummery et al., 1996, Perani et al., 1995, Chao and Martin, 2000). Furthermore, many of the brain areas showing differential activation to living and non-living stimuli are only reported in a single study.

The variability of previously reported results may be due, in part, to the statistical analysis of high-dimensional neuroimaging data. The traditional univariate statistical techniques used to analyze these data require correction for multiple comparisons to control for false positives, often making them insensitive to subtle, but widespread, effects within the brain. Therefore, univariate techniques may yield differing results depending on task-specific responses. As discussed in Section I.1.5, most current theories about the neuroanatomical basis of semantic knowledge agree that a distributed network of brain areas likely encodes this information (Patterson et al., 2007, Martin and Chao, 2001, Mitchell et al., 2008, Caramazza and Mahon, 2003). Therefore a multivariate decoding technique, which considers relationships between all features concurrently, might be able to detect distributed cortical areas that are differentially activated by living and non-living objects.

In these previous studies, due to the constraints of the imaging modality as mentioned in Section I.2, the temporal representation of these semantic categories could not be investigated. Furthermore, fMRI and PET do not directly measure neural activity, but rather a metabolic correlate. Utilizing electroencephalography (EEG) and magnetoencephalography (MEG) allows for the study of both the spatial and temporal dynamics involved in the language processing. In this chapter, we record simultaneous noninvasive EEG and MEG of healthy

participants performing a language task to explore the differences in the neural representation of living and non-living objects as well as individual words.

For successful decoding of multichannel EEG and MEG data, a classifier which is robust to high-dimensional data must be utilized. Support vector machines (SVMs) are used here to decode semantic category and individual word information from neural representations. In combination with the multichannel electro/magneto-physiological recordings performed in this chapter, SVMs allow for a multivariate examination of the spatiotemporal dynamics of the processing of words and concepts. Here, we use subject-specific decoders to study individual semantic representations, and subsequently examine the consistency between subjects and modalities using generalized SVM classifiers.

The successful decoding of semantic information from high-dimensional neural recordings not only allows for the study of language processing, but also has potential applications in the future development of language-based neuroprostheses. At the end of this chapter, we extend the SVM analysis by showing that a scalable hierarchical decoding framework that sequentially decodes word properties to narrow the search space improves on the single classifier decoding results, and may allow for the decoding of larger libraries of words and concepts.

2. METHODS

2.1. PARTICIPANTS AND DATA COLLECTION

Nine right-handed, healthy male volunteers were recorded using simultaneous scalp EEG and MEG while performing auditory and visual versions of a language task. The two tasks were performed in two separate sessions, separated by an average of 4 months. Participants

were native-English speakers between the ages of 22 to 30. This study was approved by the local institutional review board, and signed statements of consent were obtained from all subjects.

MEG was recorded using a 306-channel Elekta Neuromag Vectorview system (Stockholm, Sweden). Signals were band-pass filtered from 0.1 to 200Hz and digitized at 600Hz. Data from magnetometers and gradiometers were recorded, however only gradiometers were utilized here due to the lower noise in these sensors. Simultaneous EEG recordings were obtained from a 64-channel EEG cap at a sampling rate of 600Hz with the same filter settings as the MEG recordings. EEG was recorded using a mastoid electrode reference, but were digitally converted to a double-banana bipolar montage to reduce noise and create recording montages that were analogous to the gradiometer recordings of MEG.

2.2. LANGUAGE TASKS

A visual (SV) and an auditory version (SA) of a language task was performed by each participant. A single trial involved presentation of a written word for 300ms (in the SV task), or an auditory word 500ms in length (in the SA task), followed by a fixation point. Subjects were instructed to press a button if the presented word represented an object larger than one foot in any dimension (target trials; e.g. tiger, sofa), while refraining from responding to objects smaller than a foot (non-target trials; e.g. cricket, lipstick). Exactly half of the trials involved words representing objects larger than one foot, requiring a motor response (target trials). This required subjects to access the semantic representations of these particular words and potentially retrieve visuospatial or propositional knowledge of the associated object. Words were equally divided between living objects (animals and animal parts) and non-living objects (man-made items). Half of the trials presented a novel word which was only shown once

during the experiment while the other half of the trials presented one of ten repeated words (each shown multiple times during the experiment).

Novel words representing living and non-living objects were balanced in terms of mean number of syllables (SA: living = 1.52, non-living = 1.36, SV: living = 2.18, non-living = 2.09), letters (SA: living = 5.22, non-living = 5.21, SV: living = 6.49, non-living = 6.8), and lexical frequency (SA: living = 15.5 per million, non-living = 17.34, SV: living = 12.52, non-living = 12.45) (Francis and Kucera, 1982). These word properties were not statistically different between living and non-living object categories (Wilcoxon sign-rank, $p>0.05$). Auditory words had slightly fewer letters than visually presented words because they were required to fit within a 500ms stimulus window. Repeated words were chosen to be representative of the novel words with respect to frequency and length. Visual stimuli were presented as white text on a black background while auditory stimuli were normalized in peak volume and length. The SV and SA tasks contained unique sets of words with no overlap between the two experiments. The visual version of the task included 390 trials while the auditory version included 780 trials. Analysis of modality-specific word processing in these tasks was previously performed by Marinkovic, *et al.* (2003).

2.3. PREPROCESSING

Signals from each channel of the MEG and EEG recordings were initially bandpass filtered from 1 to 30Hz. Independent component analysis was performed on MEG and EEG signals, and EOG and EKG components were removed using a modified version of the automatic component rejection algorithm specified in Shao *et al.* (2009) using a single-class, instead of weighted, support vector machine. For each trial, the continuous recordings were epoched from 1s before to 2s after stimulus onset. Trials containing large artifacts were rejected using a

predefined amplitude threshold (300μV for EEG, 5pT/cm for MEG). Thresholds were intentionally set high to retain as much of the dataset as possible to reduce over-fitting the classifiers. After alignment to stimulus onset, waveforms from all channels were baseline corrected using a 500ms pre-stimulus period. These preprocessing steps were performed within MATLAB, using the EEGLAB 6.03b (Delorme and Makeig, 2004) and FieldTrip toolboxes (ver. 20080611, http://fieldtrip.fcdonders.nl/).

2.4. DECODING ANALYSIS

Two main components are necessary for the decoding of neural information: a feature extractor and a classifier. The goal of feature extraction is to reduce the full neural signal to a smaller number of components that are relevant for the subsequent classification task. In this case, the average amplitude in six 50ms time windows were sampled from every channel and concatenated into a large feature vector for each trial (Figure II-1). Thus, a single feature vector represents the amplitude-based spatiotemporal properties of a single trial. The six time points selected for decoding living versus non-living objects were 200, 300, 400, 500, 600, and 700ms post-stimulus, and the times selected for decoding individual words were 250, 300, 350, 400, 450, and 500ms. Previous literature suggests that the N400 component of event related potentials (ERPs) is associated with semantic processing and integration, which informed our choice of these time ranges (Bentin et al., 1993, Hagoort et al., 2004, Kutas and Hillyard, 1984, Marinkovic, 2004). These time ranges also minimize early auditory or visual effects when examining individual repeated words, and account for later activity when examining novel words (Marinkovic et al., 2003). Increasing the number of time points beyond six led to negligible increases in decoder performance, but substantially increased computational time

needed to train the classifier. Because EEG and MEG signals have different magnitudes, features were transformed to a standard normal distribution z-score ($\sim\mathcal{N}(0,1)$).

The second component of a decoding paradigm, the classifier, finds a relationship between the feature vector inputs and the corresponding word class (e.g. living/non-living object category or specific word). The generated classifier model allows for the prediction of the word class for novel data. For the classification of the MEG and EEG data recorded here, support vector machines (SVMs), as implemented by Joachims (1999), were chosen as the classifier. SVMs were chosen due to their robustness to high-dimensional data and ability to generate nonlinear decision boundaries (Vapnik, 1995). Using a set of training data from multiple classes (in our case, living and non-living object categories or individual words), SVMs attempt to find a separating boundary which maximizes the margin between these classes (Figure II-1); this reduces over-fitting and allows for good generalization when classifying novel data. For multiclass data, the SVM implementation suggested by Crammer & Singer (2002) is used.

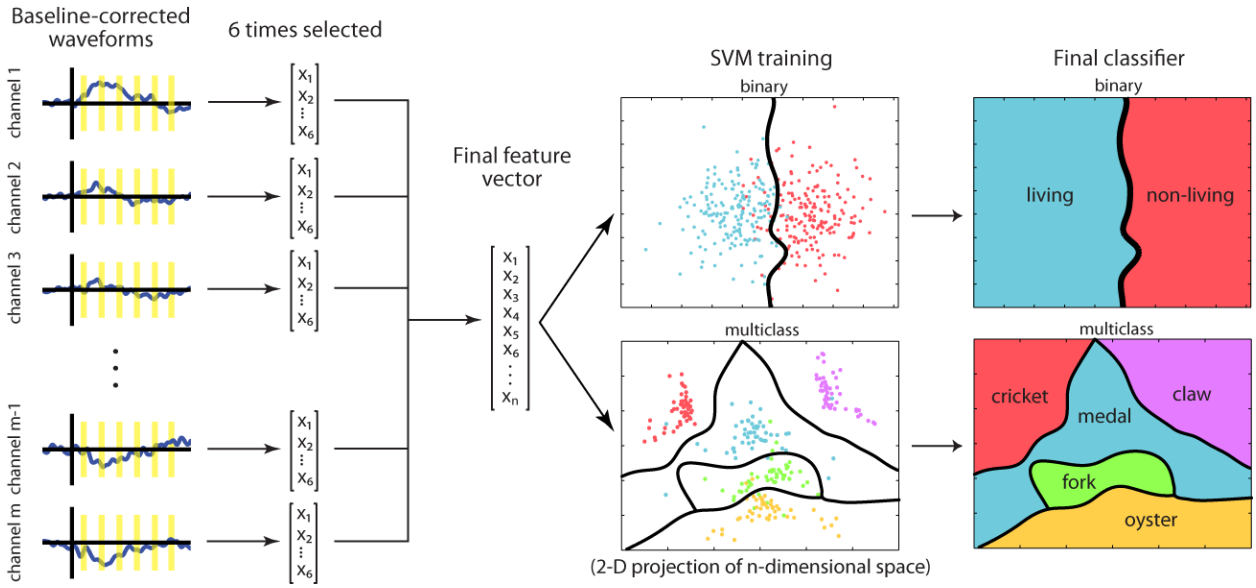

FIGURE II-1: DECODING FRAMEWORK UTILIZING AMPLITUDE-BASED FEATURE EXTRACTION AND SVMS.

The amplitude at 6 post-stimulus time points are selected from each channel and concatenated into an initial feature vector. The feature vectors from all channels are concatenated into a final feature vector. A single feature vector represents the spatio-temporal dynamics of a single trial. A nonlinear SVM is trained on these feature vectors to discriminate between the two semantic classes (living vs. non-living objects) or between individual words. This results in a decision boundary by which new trials can be classified. In the multiclass case, multiple decision boundaries are generated to separate individual classes from each other.

To estimate the accuracy of the trained classifiers, a bootstrap cross-validation was performed. This procedure splits the data into non-overlapping training and testing sets in order to evaluate the effectiveness of the classifier when encountering new data. For each round of cross-validation, one to thirty random trials of the same type (living or non-living objects in the binary case, individual words in the multiclass case) were omitted before training the SVM. The omitted (test) trials were then individually classified using the resulting model and discriminant scores averaged to generate the final classification. One thousand rounds of bootstrap cross-validation were performed to obtain an estimate of classification accuracy. Only non-repeated words were used to train the SVM to distinguish between living and non-

living objects, to allow for training on a large variety of unique stimuli within each category. By necessity, repeated words were used to train the SVM to classify individual words.

A radial-basis-function kernel, with parameter $\gamma=0.05$, was used in training the SVM to allow for nonlinear decision boundaries. The C-parameter, specifying the tradeoff between misclassification of training examples and maximal margin, was set to 1. A multiclass version of the classifier was also trained to discriminate between the five large (target) or five small (non-target) words based on the implementation in Crammer and Singer (2002). Significance thresholds, at $p=0.05$, were computed using permutation distributions generated by performing 1000 repetitions of the cross-validation procedure on trials with shuffled target labels. All subsequent results indicating statistically significant decoding accuracies utilize this metric unless otherwise specified.

2.5. DATA VISUALIZATION

The final classifier generated by the SVM consists of a set of weights that, combined with the identified support vectors, can be used to classify new trials. In the linear case, a simple weight vector is generated, the weight of each feature dictates the importance of that feature in the final classification. Thus, examining the linear SVM weights allows determination of important spatiotemporal features in the classification of living versus non-living objects. By plotting the SVM weight vector on a 2-D topographic representation of the scalp (topoplot), we can generate a map of the time-sensor points which contribute the most to the final classifier.

In the individual word (multiclass) case, a set of weights is generated for each of the five words. Highly variable weights associated with a particular feature indicate that the classifier more heavily utilizes this feature to discriminate different words. Thus, the variance of the

weights for each feature was computed as a metric of relative importance of each time-sensor point and plotted in a topoplot. Because the lead field of a MEG planar gradiometer is directly under the sensors, these topoplots are a valid way of exploring the cortical areas contributing to the discrimination (Hämäläinen and Ilmoniemi, 1994).

Confusion matrices were also generated in the individual word (multiclass) case to compare the actual word presented with the word predicted by the classifier. These matrices indicate the type and quantity of errors generated by the multiclass SVM when decoding individual words and allow for a systematic analysis of the words that were most difficult to classify. Any given row of these matrices shows the distribution of classification of a particular word with the diagonal indicating correct classification and off-diagonals indicating errors.

2.6. INTERMODALITY AND INTERSUBJECT DECODING

To study supramodal representations, we trained an SVM on features from either the auditory or visual tasks from a single subject. Data from the same subject, but opposing modality, was then used to test the classifier. Because no word overlap was present between the repeated words in the SV and SA versions of the task, this could only be done for the classification of categories (animals versus non-living objects).

Intersubject decoding was also performed to examine the consistency of language-related representations between individuals. In this case, an SVM was trained on data from all but one subject within a single modality. The data from the remaining subject was then used as test data. This was repeated by omitting each subject in turn. This analysis was performed on both living/non-living category and individual words.

2.7. HIERARCHICAL TREE DECODING

Although utilizing a single multiclass decoder to distinguish individual word representations often works well, it does not directly incorporate *a priori* knowledge about semantic classes and the features which best discriminate these categories. Moreover, as the number of classes or features grow, it becomes more difficult for a single classifier to perform well with a fixed size of training data. To combine information from the classifier models generated to decode semantic category and individual words, we implemented a hierarchical framework which attempts to decode word properties sequentially. Given an unknown word, the tree decoder first classifies it as either a large (target) or small (non-target) object. The word is then classified as living or non-living object, and finally as an individual word within the predicted semantic category. This allows the appropriate features to be used to decode each word property, narrowing the search space before individual words are decoded. Such a construct is easily scalable and could allow for the eventual decoding of larger libraries of words.

As a proof-of-concept, a 3-level hierarchical decoding construct was implemented using a set of SVMs for each level of the tree. Amplitude (six time points from 200-700ms) and spectral features (8-12Hz power at six time points from 200-700ms) were first utilized to decode whether an unknown word represented a large (target) or small (non-target) object. The spectral features allowed for motor intent to contribute to this initial classification. In the second and third levels of the tree, amplitude features from 200-700ms were utilized to decode living/non-living object category, and amplitude features from 250-500ms were used to decode individual words. A separate SVM was trained (using the same parameters described in section 2.4) for each level of the tree using the appropriate trials. Upon classification of a new trial, the result of earlier levels determined which of the trained models would be utilized to

decode subsequent features. For example, if a novel trial was first decoded as a large object (target trial), and subsequently decoded as a living object, the final classifier would label the trial as either a “dinosaur”, “python” or “steer”.

To compare performance to a single multiclass decoder, an SVM was trained to discriminate between all 10 words using the full set of amplitude and spectral features used in the hierarchical tree decoding. A bootstrap cross-validation with 1000 repetitions was again used to estimate the accuracy of this decoder.

3. RESULTS

3.1. BEHAVIORAL RESULTS

To ensure that behavioral responses to different trial types did not contribute to the decoding of words and semantic categories, we first analyzed the accuracy and response times of button presses (to large objects) for all subjects. Accuracy of behavioral responses ranged from 71.6 to 95.5% with a mean of 90.3±1.4% across subjects. Mean response times varied from 760 to 1152ms with a cross-subject mean of 943±27ms. Mean accuracies for living and nonliving object categories across subjects were 90.4±1.6% and 90.2±1.6% respectively. Mean response times for living and nonliving object categories were 947±30ms and 962±25ms. Accuracies and response times were not significantly different between living and nonliving object trials for any of the subjects (Wilcoxon sign-rank, $p>0.05$). It is therefore unlikely that differential behavioral responses influenced subsequent decoding analyses. Accuracies were not significantly different between SV and SA tasks (Wilcoxon, $p>0.05$), although mean response times were shorter for the visual task (SV: 864ms, SA: 1023ms, Wilcoxon, $p<0.00001$). As expected, response times were shorter for repeated versus novel words

(repeated: 868ms, novel: 1023ms, Wilcoxon, $p<0.001$). Mean accuracies and times were not significantly different between individual repeated words for any subject (ANOVA, $p>0.05$).

3.2. DECODING OF SEMANTIC CATEGORY

We first attempted to train a SVM to decode living versus non-living objects. The SVM was trained separately on EEG features, MEG features, and both combined. Figure II-2A-B illustrates the decode accuracies after averaging 5 trials (chance accuracy = 50%). When utilizing EEG features alone, data from 7 of the 9 subjects in the SV task and 6 of 9 in the SA task, showed statistically significant decoding accuracy (permutation test, $p<0.05$). When utilizing MEG features alone, data from 8 of the 9 subjects in SV and 7 of 9 in SA showed significant decoding accuracy (permutation test, $p<0.05$). Statistically significant decoding accuracy was obtained from all subjects when utilizing combined EEG and MEG features in both SV and SA tasks. When utilizing combined EEG and MEG features, accuracies ranged from 63-86% (mean±s.e. = 76±2%) for the SV task and 62-91% (mean±s.e. = 75±3%) for the SA task. Training on both MEG and EEG features increased accuracies by an average of 12% for the SV task and 10% for the SA task over using EEG features alone and 8% (SV) and 4% (SA) over MEG features alone (Wilcoxon sign-rank, $p<0.05$). Accuracies for the SV and SA task were not statistically different in any set of features when discriminating between living and non-living objects (Wilcoxon, $p>0.05$). These results suggest that high-dimensional machine-learning algorithms, such as SVMs, are able to robustly extract semantic category information from multichannel electro/magneto-physiological recordings.

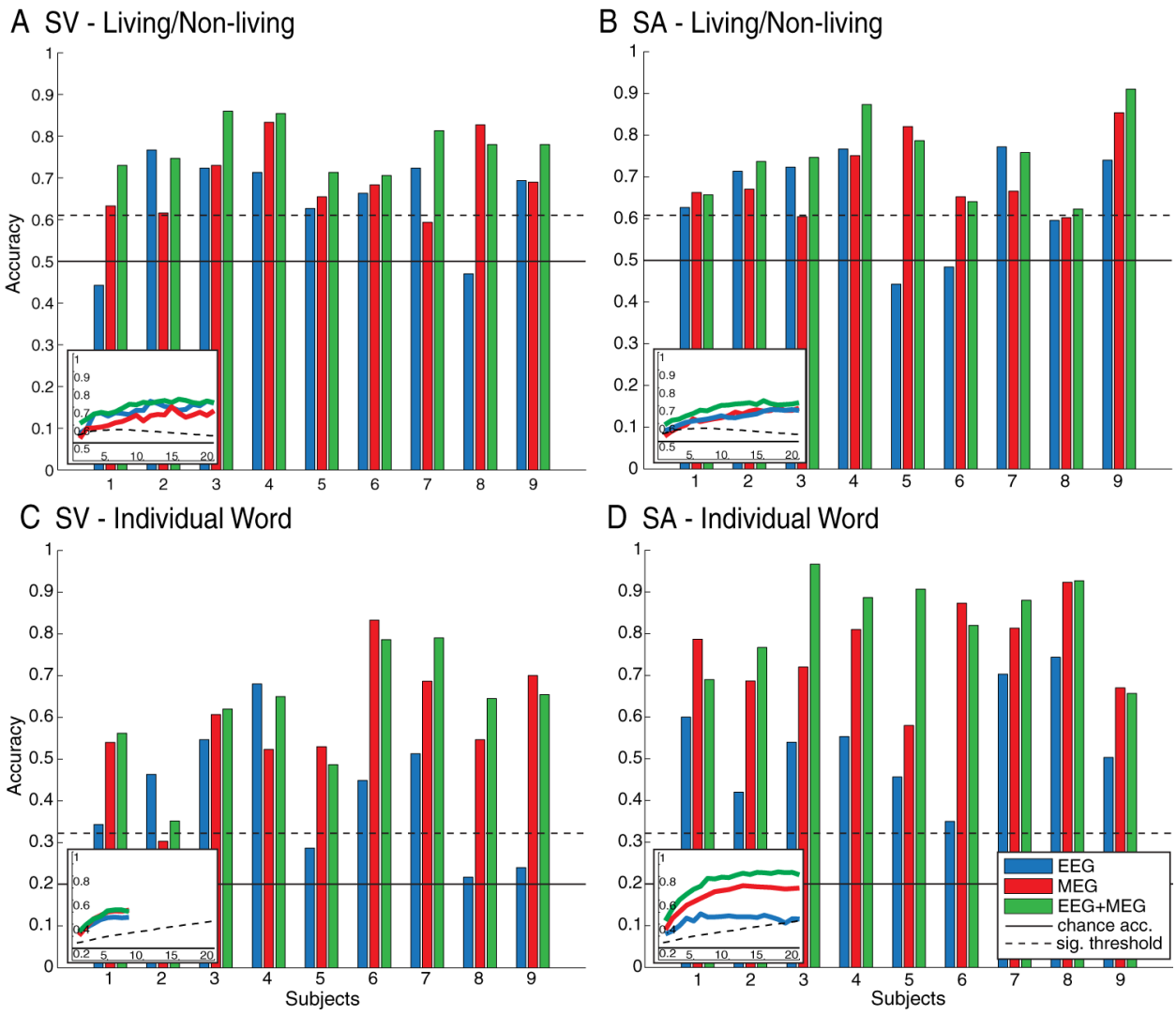

FIGURE II-2: DECODING ACCURACY WHEN DISTINGUISHING BETWEEN LIVING AND NON-LIVING OBJECTS OR INDIVIDUAL WORDS.

The bar graphs illustrate classifier accuracy for each subject when distinguishing between living and non-living object category (A-B) or between individual words (C-D) after averaging 5 trials. Inset panels illustrate mean decoding accuracy as a function of the number of trials averaged. Blue indicates the use of EEG features, red indicates MEG features, and green indicates that both EEG and MEG features were used. In both the main figure and insets, chance accuracy (0.5 for living/non-living and 0.2 for individual words) is shown as the horizontal black line and accuracies above the dashed line are statistically significant (permutation test, $p<0.05$). A-B) Data from all subjects show significant decoding ability in at least one set of features. In all cases, utilizing combined EEG and MEG features resulted in significant decode accuracies. C-D) When utilizing both EEG and MEG features, decoding performance when distinguishing individual words is statistically significant in all cases, and exceeds 95% accuracy in the SA task.

To explore the effect of the number of trials averaged on decoding accuracy, we also performed a leave-n-out cross-validation on all sets of features with all subjects (Figure II-2 inset panels). Not surprisingly, increasing the number of trials averaged resulted in increased decode performance in all cases. However, averaging more than approximately 7 trials resulted in only marginal additional increases in performance.

3.3. DECODING OF INDIVIDUAL WORD REPRESENTATIONS

We subsequently examined SVM decoding of individual word representations utilizing multiclass SVMs. We trained and tested classifiers on either the 5 repeated non-target (small objects) or target words (large objects) to decode individual word representations without the potential motor confound (chance accuracy = 20%). The requirement for a motor action (button-press when the presented object was larger than one foot) may result in the decoding of that volitional response, rather than word processing information *per se*, when examining differences between all 10 words. The ability of the classifier to predict the observed word was statistically significant for all subjects after averaging 5 trials in at least one set of features (permutation test, $p<0.05$) (Figure II-2C-D). Accuracies varied from 32-79% (mean±s.e. = 60±5%) using combined EEG/MEG features for the SV task (chance accuracy is 20%). For the auditory task, accuracies varied from 66-97% (mean±s.e. = 83±4%). Training the SVM classifier on both EEG and MEG features increased average decode performance by 18% for the SV task and 29% for the SA task over using EEG features alone and 2% (SV) and 7% (SA) over MEG features alone (Wilcoxon, $p<0.05$). The decoding accuracies of the SV and SA tasks when utilizing solely EEG features were not significantly different (Wilcoxon, $p>0.05$). However, utilizing MEG alone or both feature types resulted in significantly better performance in the SA data than utilizing the corresponding feature sets in the SV data (Wilcoxon, $p<0.01$).

The SA task contained twice as many trials as the SV task (780 for SA versus 390 for SV) which may have resulted in the difference in decoding accuracy between the two presentation modalities. By utilizing only the first 390 trials of the SA task, accuracy of the multiclass decoder after averaging 5 trials (mean±s.e. = 61±4%) was not significantly different from SV performance (mean±s.e. = 60±5%) (Wilcoxon, $p>0.05$).

Again, increasing the number of trials averaged increases decoding performance substantially (Figure II-2C-D inset panels). In the case of individual word decoding for the SV task, there is a slight decrease in accuracy when the number of trials averaged is increased from 6 to 8. This is likely due to the fact that increasing the number of trials averaged causes a corresponding decrease in the number of trials used for training the SVM, leading to a less robust classifier. This is especially pronounced in the multiclass SV case on account of the relatively smaller number of total trials per condition when compared to the SA task. These data also illustrate that combining EEG and MEG features improves accuracy over either feature set alone. Taken together, these results demonstrate surprisingly robust ability to decode individual words from spatiotemporal features computed from multichannel electrophysiology.

3.4. DECODING USING PROBABILISTIC CLASSIFIERS

While a decoding analysis is a powerful method for exploring electro/magneto-physiological data, not all classification algorithms are suited for such an analysis. To demonstrate the advantages of utilizing machine-learning techniques robust to high-dimensional data, we compared the decode accuracy obtained when using SVMs (sections 3.2 and 3.3) to the use of a popular probabilistic classifier. Because traditional Fisher linear discriminant analysis and Bayesian decoders are unable to handle cases in which the number

of features is close to, or exceeds, the number of trials, we utilized a naïve Bayes classifier. Naïve Bayes classifiers assume independence of features, and are thus able to train and classify this particular set of MEG/EEG features.

When classifying living/non-living category using MEG and EEG features, a naïve Bayes classifier resulted in average accuracies of 54±4% and 51±3% for SV and SA respectively (chance=50%). This was significantly lower than the SVM classification of the same data (76% for SV, 75% for SA, Wilcoxon sign-rank, $p<0.005$), and in fact not statistically different from chance. Similarly, when classifying individual words using MEG and EEG features, a naïve Bayes classifier yielded accuracies of 41±4% and 46±3% for SV and SA data respectively (chance=20%). This, again, is significantly lower than the classification using an SVM (60% for SV, 83% for SA, $p<0.005$). These results suggest that a decoding analysis of MEG/EEG data requires techniques which are robust to high-dimensional data. In this case, SVMs, when compared to a naïve Bayes classifier, are better able to handle such data and can provide insight into the spatiotemporal representations of semantic knowledge.

3.5. SVM WEIGHTS DEMONSTRATE SPATIOTEMPORAL DYNAMICS

Examining the SVM weights allows us to determine the features which were most important in the generation of the final SVM classifier (Figure II-3). In the linear case, the weight of each feature dictates the importance of that feature in the final classification. Because the weights of a nonlinear classifier cannot be easily visualized, we utilized linear SVMs when examining classifier weights. The performance when using nonlinear SVMs was greater than the performance of the linear SVMs by 3.3% on average (Wilcoxon sign-rank, $p<0.05$), however decode accuracy remained high in the linear case. In all cases where the nonlinear SVM yielded statistically significant decode accuracy, the linear SVM also yielded

statistically significant accuracy. Thus, examining the linear SVM weights allows determination of important spatiotemporal features in the classification.

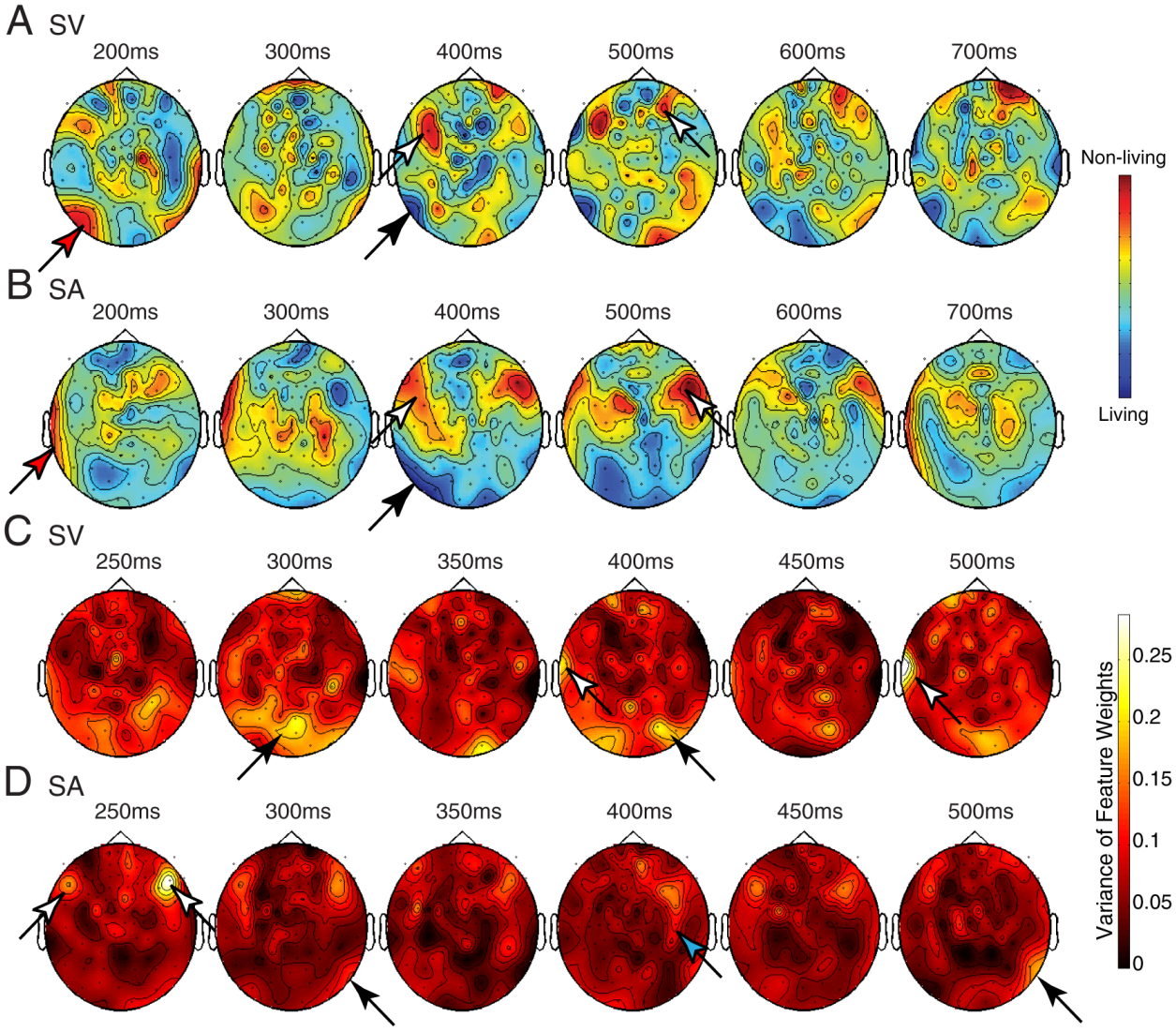

FIGURE II-3: CLASSIFIER WEIGHTS SHOW IMPORTANT TIMES AND LOCATIONS FOR DECODING.

A-B) SVM weights of the classifier trained on living and non-living object categories in MEG show areas of significant differences. Areas of dark red indicate biases in classification towards non-living objects, and blue denotes biases towards animals and living objects. Averaged weights across all subjects are shown at each sensor-time point for the (A) visual and (B) auditory tasks. Bilateral anterior temporal and inferior frontal differences are seen at 400-600ms during both the SV and SA tasks (white arrows). Left temporal-occipital differences showing larger responses to objects are apparent at 200ms (red arrows) with differences showing larger responses to living objects occurring at 400-700ms in both modalities (black arrows). C-D) Variance of SVM weights is shown at each time-sensor point indicating relative importance of each feature in individual word

(Figure II-3 continued) discrimination. Features with larger variance indicate larger separation between the SVM weights in that particular dimension and correlate with increased discrimination ability. C) Extracranial weights from the SV task indicate occipital significance around 300-400ms (black arrows) and inferior temporal significance at several times (white arrows). D) Weights from the SA task show bilateral anterior temporal and inferior frontal significance from 250-450ms (white arrows), and inferior occipital significance at 300 and 500ms (black arrow). Inferior parietal significance is also seen from 350-400ms (blue arrow).

Averaged weights across subjects for the visual (Figure II-3A) and auditory (Figure II-3B) tasks show a broadly distributed pattern of information-specific activity. Large weights are observed at all sampled time points and across both hemispheres. In particular, bilateral anterior temporal and inferior frontal weights increase to inanimate objects relative to living objects from 400-600ms. A concurrent increase of SVM weights in response to living over non-living objects is present at left inferior temporal-occipital sensors from 400-700ms. Interestingly, an early temporal-occipital increase in weights to non-living objects is seen at an earlier latency of 200ms. While left inferior temporal-occipital activation to animals has previously been observed, the earlier activation to non-living objects has not been reported.

When decoding individual word representations, the multiclass SVM generates one set of weights for each class. For visualization purposes, the variance of the SVM weights across words for each time-sensor point was computed and displayed (Figure II-3C-D). Features with higher variances differ more across classes, generally making them more important in the final classification. These data also show fairly distributed set of time-sensor points which contribute to the decoding. The SV data showed inferior occipital increase in weight variance from 300-400ms, and inferior temporal activation from 400-500ms (Figure II-3C). The SA task showed increased weight variance in bilateral anterior temporal areas from 250-450ms with increases in posterior sensors at 300 and 500ms (Figure II-3D).

To further study the time-course of semantic information present in MEG signals, a cumulative window decoding analysis was performed to decode individual word identity from these neural recordings. MEG amplitude features were computed in 30ms windows for the repeated words and the SVM decoder attempted to predict individual word identity from a cumulative set of windows beginning at 0ms. Data from both SA and SV tasks demonstrate significant decode accuracies beginning at approximately 150ms and saturating around 500ms (Figure II-4). This is roughly consistent with the time course of the SVM weights observed in Figure II-3, suggesting that word-specific information occurs as early as 150-200ms.

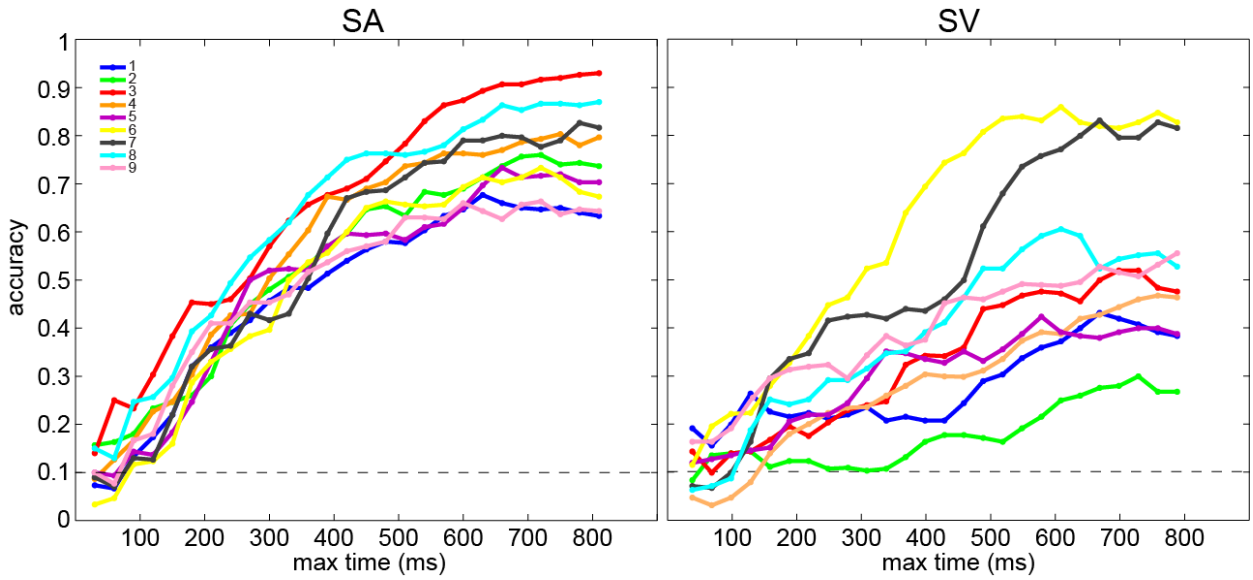

FIGURE II-4: CUMULATIVE TIME DECODING DEMONSTRATES TIME COURSE OF INFORMATION

Decoding the ten repeated words in the SA (left) and SV (right) tasks using MEG data from cumulative 50ms time windows shows the time course of information. In both cases, information content provides significant decoding accuracy starting at approximately 150ms and this accuracy saturates at approximately 600ms for both modalities.

3.6. SEMANTIC CATEGORY INFORMATION IS SPATIALLY DISTRIBUTED

While SVM weights demonstrate that a large number of sensors at various time points contribute to the classification of semantic category, it is important to understand whether the

accuracy achieved in this decoding is a result of the information present in a small number of electrodes. To determine whether this was the case, a SVM was trained and tested on the same set amplitude features in the 6 time windows used for the other analyses, but instead of utilizing all sensors, decoding was performed on single sensors (Figure II-5). The mean accuracy for this single sensor decoding was 52±2%, and ranged from 45-60%. To test whether any of these accuracies were statistically significant, a null distribution was generated by shuffling the labels of trials. Only 3.7% of the channels demonstrated statistically significant accuracy ($p<0.05$), yet no single channel approached the mean accuracy when using all channels of 71%. The skewness of the distribution of accuracies was computed as 0.67 (with the skewness of the null distribution at -0.02), showing that the distribution was skewed towards above-chance accuracies. This may suggest that a large number of channels had accuracies above 50% and carried some semantic information, despite not being statistically significant. This is consistent with the idea that the representation of semantic category is highly distributed, and that no single sensor is able to provide the same amount of information that can be extracted when all sensors are combined.

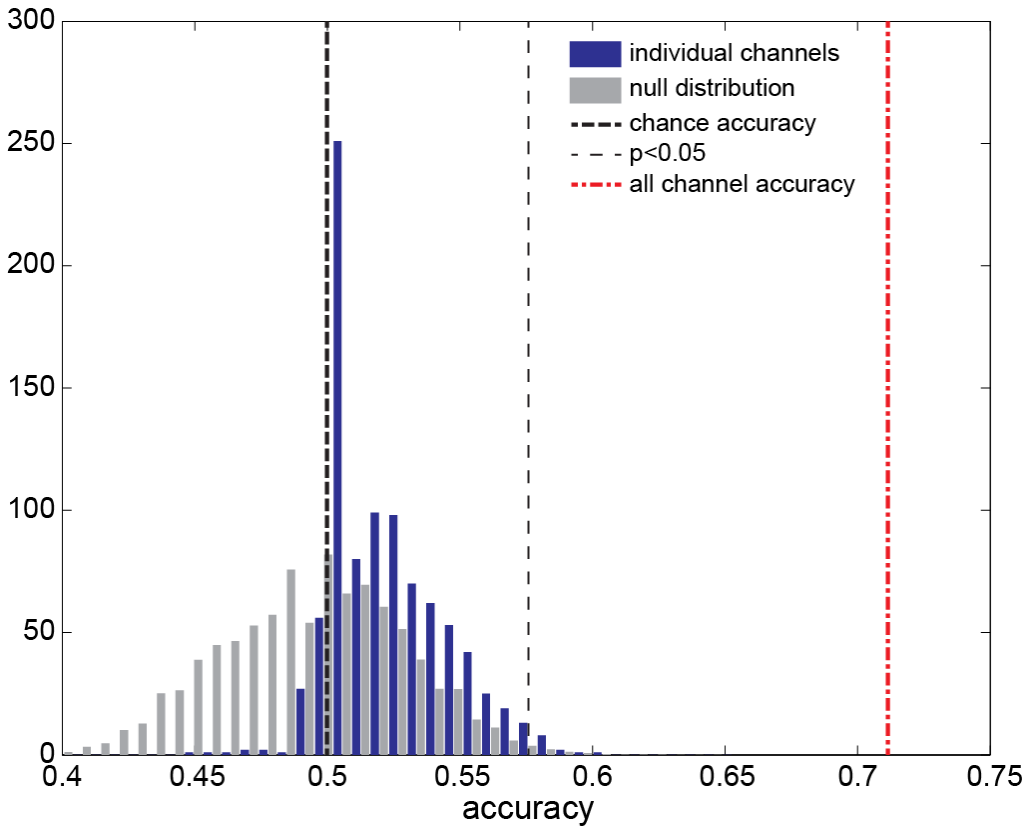

FIGURE II-5: DECODING OF SINGLE CHANNELS DEMONSTRATES THE SPATIALLY DISTRIBUTED NATURE OF SEMANTIC CATEGORY INFORMATION

The decoding of single channels failed to yield accuracies near the accuracy obtained when using all channels of 71%. However, the distribution of accuracies was highly skewed (skewness = 0.67), suggesting that a large number of channels with accuracies above 50% contained semantic information despite the fact that only 3.7% of channels demonstrated significant accuracy with regard to the computed null distribution ($p<0.05$).

3.7. SYSTEMATIC ERRORS IN INDIVIDUAL WORD DECODING REVEAL SEMANTIC STRUCTURE

Confusion matrices were constructed to analyze errors generated when discriminating between all 10 repeated words (Figure II-6A). The actual stimulus words are present along the vertical axis while the words predicted by the classifier are present along the horizontal axis. The colors along any given row (*actual word*) indicate the proportion of trials of that word which were classified as each of the possible choices (*predicted words*) (i.e. the confusion rate). Therefore, if the classifier correctly classified the word "feather" in all cases, the first element in

the row corresponding to “feather” would be 1 (i.e. “feather” was always classified as “feather”) with all other elements being 0 (i.e. “feather” was never classified as any other word). Therefore, the diagonal elements in the matrix display correctly classified trials.

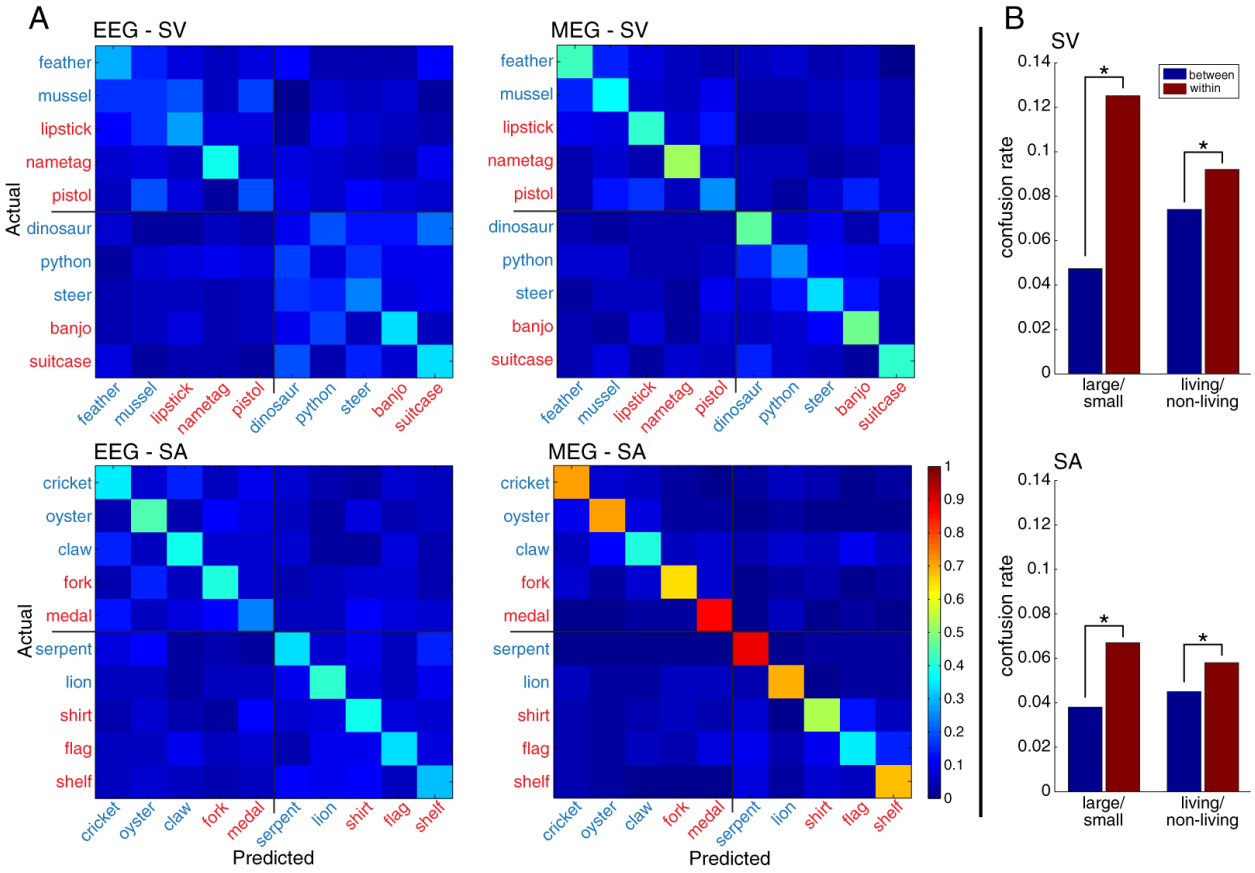

FIGURE II-6: INDIVIDUAL WORD DECODING CONFUSION MATRICES.

A) Averaged confusion matrices for decoding all 10 individual words (averaging 5 trials) indicate the types of errors made. The vertical axis displays the actual stimulus word while the horizontal axis displays the word predicted by the classifier. The colors along any given row (*actual word*) indicate the proportion of trials of that word which were classified as each of the possible choices (*predicted words*). The diagonal elements display correctly classified trials. Words are sorted into small and large objects (divided by black lines), and living or non-living categories (blue and red text). These matrices demonstrate a significant ability to decode individual words without regard to large/small conditions. B) Within and between-category confusion rates are shown for the large/small and living/non-living object distinctions. In all cases, confusion rates between categories are statistically lower than confusion rates within each category.

Visual examination of confusion matrices confirms that decoding of the MEG auditory data yields the highest accuracy, followed by EEG auditory data, followed by data from the

visual task. The confusion matrices of combined EEG and MEG data were virtually identical to the confusion matrices generated to MEG data alone. A larger confusion rate is visually apparent within target (large object) or non-target (small object) classes (upper left and lower right corners), compared to between the two classes (lower left and upper right). The required motor response associated with the target trials may be providing additional non-language information allowing for a decreased error rate when decoding between all 10 repeated words (as discussed in section 3.3). Despite this, the ability to decode individual words is seen within the large and small object groups; providing additional evidence that word-specific information is present in the neural signals being classified.

To quantify the effects of semantic category and large versus small objects on confusion rates, we performed a 3-way ANOVA on these data (Figure II-6B). This was performed to determine if two words which were within the same class (e.g. both living objects, both small objects, etc.) had a higher confusion rate than two words in different classes. In other words, the ANOVA compares differences in "within-class" confusion rates to "between-class" confusion rates. The ANOVA analysis involved three factors (living/non-living, large/small, and subjects) with two levels in the categorical factors (within-class or between class) and 9 levels in the subject factor (one for each subject).

For the SV task, the average large/small between-class confusion rate (mean±s.e. = 0.0472±0.027) was significantly smaller than large/small within-class confusion (0.125±0.045; $F=45.72$, $p<0.00001$). Average living/non-living object between-class confusion (0.074±0.037) was significantly smaller than living/non-living object within-class confusion (0.092±0.043; $F=8.59$, $p<0.005$). For the SA task, the average large/small between-class confusion (0.038±0.028) was significantly smaller than large/small within-class confusion (0.067±0.036;

F=20.28, p<0.00001). Average living/non-living object between-class confusion (0.045±0.031) was also significantly smaller than living/non-living object within-class confusion (0.058±0.034; F=7.99, p<0.05). This shows that it is more difficult for the classifier to discriminate words within the same semantic category than words of different categories. This suggests semantically related words have similar neural representations, and provides further evidence of the natural distinction between living and non-living objects.

3.8. LOW-LEVEL STIMULUS PROPERTIES

It is possible that the generated classifiers are utilizing neural activity related to low-level visual or auditory stimulus properties when decoding individual words. For example, the classifier may be decoding brain activity which is specific for the number of letters in the visual word or the number of syllables in the acoustic word, and not the semantic information associated with the word. To test this, we performed a shuffling based on stimulus properties to evaluate this potential confounding factor. Within either the 5 target or non-target words, we randomly swapped half of the trials between two words with equal numbers of letters or numbers of syllables, thus creating two categories with consistent sensory characteristics but scrambled lexical referents, while leaving the remaining three words unchanged. If the decode ability was solely based on either of these visual or phonetic properties of the stimulus, we would see no change in accuracy. In fact, the decoding accuracy of these sensory based categories dropped by 24% (letters) and 30% (syllables) (Wilcoxon sign-rank, p<0.01). Accuracies remained statistically above chance due to the fact that trials associated with 3 of the 5 words were left unchanged.

Although these low-level properties were not solely responsible for the decode ability, if these stimulus characteristics contributed information to the decoding, shuffling trials between

two words with different sensory characteristics would result in a larger drop in accuracy compared to shuffling between words with consistent sensory characteristics. The drop in performance when swapping trials between words with similar sensory characteristics was not significantly different from the performance when swapping trials between words with different sensory characteristics (25% for letters and 28% for syllables, Wilcoxon, $p>0.05$). These sensory characteristics therefore did not contribute significantly to the decoding of individual words in the visual version of the task.

We performed the same shuffling analysis for the SA task as well. The drop in performance was 23% when shuffling between words with the same number of syllables (Wilcoxon, $p<0.01$). This decrease in accuracy was not statistically different from the decrease in accuracy when shuffling between words with different numbers of syllables (20%, Wilcoxon, $p>0.05$).

To control for the possibility of frequency-related acoustic properties of the words affecting the decode analysis (in the SA task), we attempted to predict stimulus properties using the same set of neural features used in the individual word decoding. In this case, the SVM algorithm performed a regression instead of classification to predict the power of the acoustic stimuli within five frequency bands (250-500Hz, 500Hz-1kHz, 1-2kHz, 2-4kHz, and 4-8kHz). If any of these acoustic properties contribute to the decoding of individual words, we would expect that a SVM trained on the previously used features would also be able to predict the power in these auditory frequency bands. To statistically test these results, a permutation distribution was computed by shuffling trials so that each trial was associated with a random set of stimulus band-power values for 2000 trainings of the SVM regression. The root-mean-squared error was computed for each of these repetitions, resulting in a distribution of errors

for the case that no information about stimulus band-power was present in the computed features. The root-mean-square error of this regression was not statistically significant based on a permutation distribution computed by shuffling the stimuli ($p>0.05$, Figure II-7). This result suggests that the decoding of individual words was not solely a result of differential representation of low-level properties of the auditory stimulus such as acoustic power.

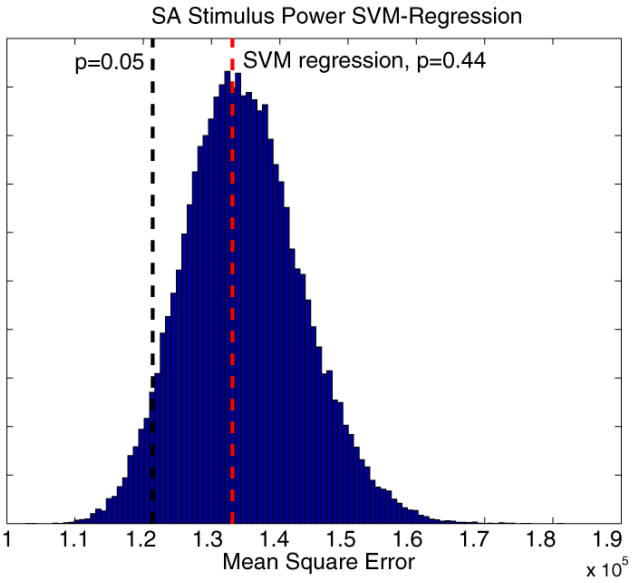

FIGURE II-7: AUDITORY STIMULUS SVM REGRESSION.

Utilizing the same features and SVM algorithm as in the classification of individual words, the SVM was run in regression mode to attempt prediction of auditory stimulus power in 5 frequencies bands. The null distribution was computed by shuffling trials 2000 times. Obtaining a p-value of less than 0.05 required the mean squared error of the regression to be less than 1.21×10^5. However, the resulting error was 1.33×10^5 with a corresponding p-value of 0.44. This is not statistically significant and suggests that a basic stimulus property such as band-power did not contribute to the classification of individual words.

3.9. INTER-SUBJECT AND INTER-MODALITY DECODING

To investigate supramodal contributions to the generated classifiers, SVMs were trained on one stimulus modality and tested on the other modality. When training on visual data and testing on auditory data, statistically significant decode accuracies was obtained in 3 of 9 subjects (Figure II-8A) with a mean accuracy across all subjects of 57.5±3.0%. When training

on the auditory modality and testing on the visual modality, data from 5 of 9 subjects showed significant decode accuracies with a mean accuracy across all subjects of 67.7±4.1%. This suggests that the models generated with features from either version of the task contain supramodal semantic information. This is more apparent in the case where the training set was larger and better able to produce a robust classifier (training on SA, testing on SV). By increasing the number of trials averaged, performance improves, as seen previously (Figure II-9).

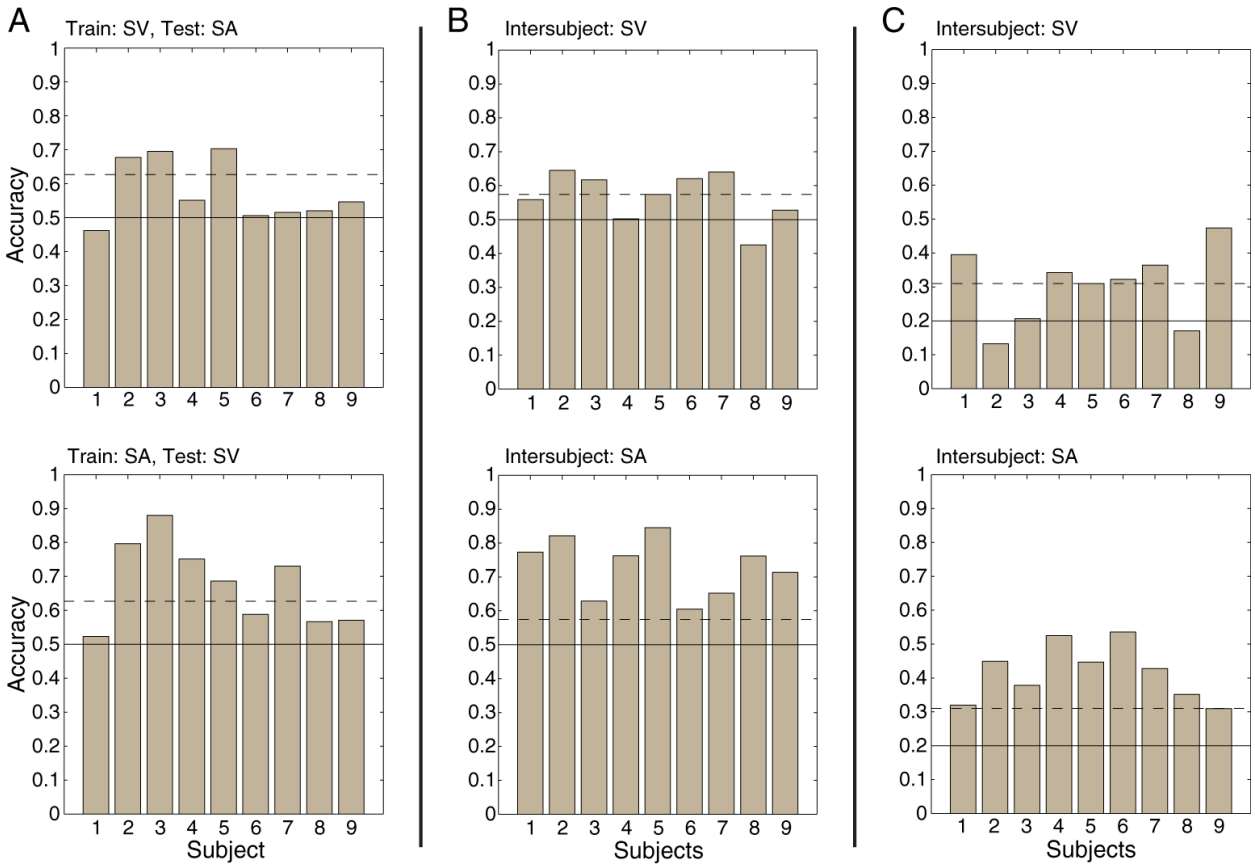

FIGURE II-8: INTERMODALITY AND INTERSUBJECT CLASSIFICATION SHOWS WORD AND CATEGORY REPRESENTATION CONSISTENCIES.

Accuracies of intermodality and intersubject decoding are shown after averaging 10 trials. Chance accuracy is indicated by solid horizontal line and the statistically significant threshold is shown by dashed line (permutation test, $p<0.05$). A) Training on living/non-living object data from SV and testing on data from SA results in data from 3 of 9 subjects showing statistically significant decode ability while training on SA and testing on SV results in data from 5 of 9 subjects showing significant decode ability. This indicates supramodal semantic information is encoded within the

(Figure II-8 continued) classification models generated by the SVM. B) Training an SVM on living/non-living object data from all but one subject and testing on the final subject results in data from 5 of 9 and 9 of 9 subjects showing statistically significant decode within the SV and SA modalities respectively. C) Training an SVM on individual word representations from all but one subject and testing on the final subject results in data from 6 of 9 and 9 of 9 subjects showing statistically significant decode ability. This indicates intersubject consistency in the neural representation of these words.

We also investigated the ability to train a generalized, subject-nonspecific decoder by training an SVM on data from all but one subject, and testing on the final subject's data. The accuracy obtained from such a cross-validation is an indication of the consistency of language-related representations between individuals. In the first case, an SVM was trained to discriminate between living and non-living object categories. Data from 5 of 9 subjects for SV and all subjects for SA showed statistically significant decoding performance (Figure II-8B, $p<0.05$). Mean accuracies were 56.8±2.4% and 72.9±2.8% for SV and SA respectively.

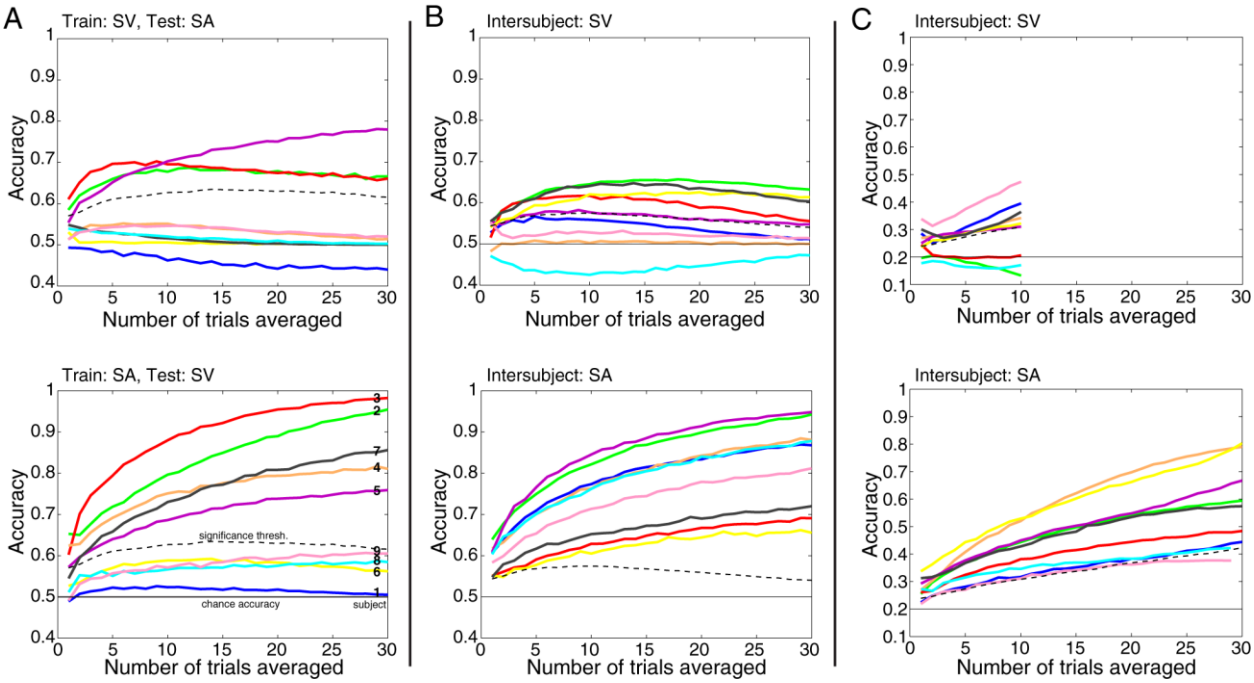

FIGURE II-9: INTERSUBJECT AND INTERMODAL DECODE PERFORMANCE WITH INCREASING NUMBERS OF TRIALS

Plots illustrate decoding accuracies per subject as the number of trials averaged increases. Horizontal line illustrates chance accuracy with dashed line showing significance threshold

(Figure II-9 continued) determined via computation of a permutation distribution ($p<0.05$). A) Training on animal/object data from SV and testing on data from SA results in 3 of 9 subjects' data showing statistically significant decode ability while training on SA and testing on SV results in 5 of 9 subjects' data showing significant decode ability at 10 trials averaged. This indicates supramodal semantic information is encoded within the classification models generated by the SVM. B) Training an SVM on animal/object data from all but one subject and testing on the final subject results in data from 5 of 9 and 9 of 9 subjects showing statistically significant decode within the SV and SA modalities respectively at 10 averaged trials. This indicates individual word representations are fairly consistent between subjects. C) Training an SVM on individual word representations from all but one subject and testing on the final subject results in data from 6 of 9 and 9 of 9 subjects showing statistically significant decode ability at 10 averaged trials.

A generalized SVM was also trained to discriminate between 5 large or small repeated words. Figure II-8C indicates that in 6 of 9 cases for SV and all cases for SA, the decoding accuracy was significantly above chance levels. Mean accuracies were 30.2±3.7% for SV and 41.3±2.7% for SA (chance = 20%). Despite the fact that MEG sensor positions are variable between subjects, above-chance accuracies were obtained, suggesting that some word-specific information is consistent between individuals. Not surprisingly, however, subject-specific classifiers still yield significantly higher decode accuracies.

3.10. EXTENSIBLE HIERARCHICAL FRAMEWORK

To explore the potential practical use of machine-learning algorithms to decode larger libraries of words, we used SVM classifiers within the larger construct of a hierarchical framework (Figure II-10). Such a paradigm is easily scalable and may allow for the eventual decoding of a large number of individual words or concepts. Utilizing a hierarchical construct allows for the incorporation of *a priori* knowledge about semantic classes and the features which best discriminate these categories.

The average accuracy of all branches of the tree for the SA task was over 80% and accuracies at each level of the decoder were above 80% for all but 2 subjects (Figure II-10A-B). By examining cumulative accuracies at each level of the tree, we find that errors propagate

from earlier levels, as expected, but accuracy ultimately remain above 60% in all cases (Figure II-10C). The mean overall accuracy of the tree decoder was 70%, significantly higher than the 67% accuracy of a single multiclass SVM trained on all 10 words (Wilcoxon sign-rank, $p<0.05$) (Figure II-10D). Data from all subjects, but subject 7, showed an improvement over the single SVM classifier when using the tree decoder. Thus, the hierarchical tree framework, by incorporating *a priori* knowledge of semantic properties, and utilizing specific features at each level, allows representations of individual word properties to be decoded more accurately than using a single multiclass decoder which treats each word as an independent entity.

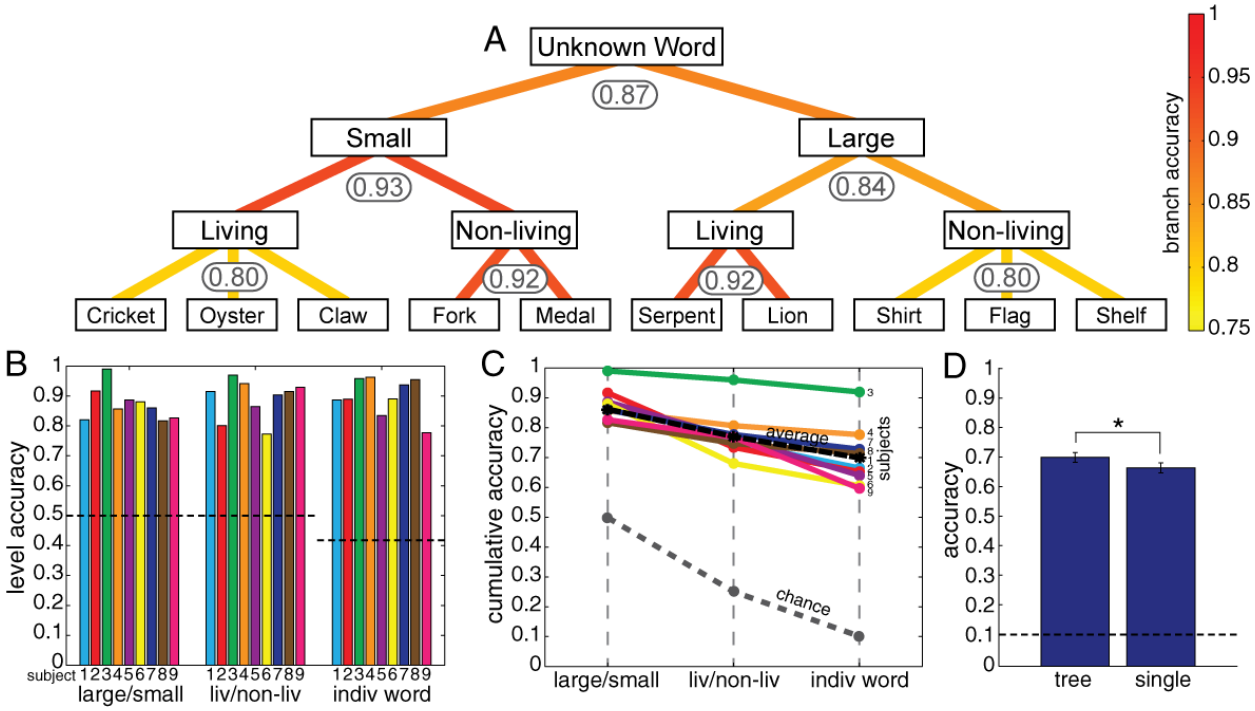

FIGURE II-10: HIERARCHICAL DECODING IMPROVES CLASSIFICATION PERFORMANCE.

A three-level hierarchical tree decoder was utilized to first decode the large/small distinction (utilizing amplitude and spectral features), then the living/non-living object category (utilizing 200-700ms amplitude features), and finally the individual word (utilizing 250-500ms amplitude features). Data from decoding of the SA task are shown. A) Average accuracies at each branch of the tree are shown with corresponding colors. Accuracies remain above 80% for all branches. B) Accuracies at each level of the decoder are shown on a per subject basis with dotted lines indicating chance accuracy. C) Cumulative accuracies at each level decrease as errors propagate through levels of the tree, but remain above 60%. D) Performance of the hierarchical tree is a significant

(Figure II-10 continued) improvement (Wilcoxon sign-rank, $p<0.05$) over training a single multi-class SVM to discriminate between all 10 words.

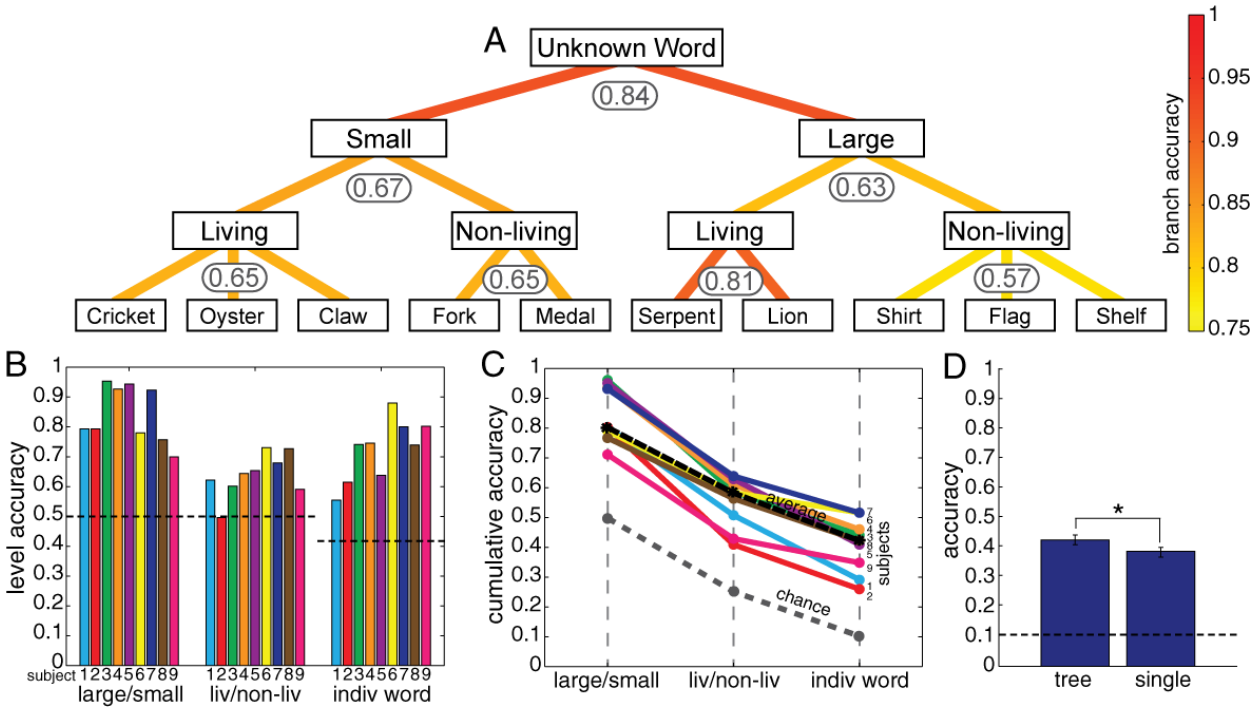

FIGURE II-11: HIERARCHICAL DECODING FOR SV TASK.

This figure is analogous to figure 6 in the main text generated using data from the SV task. A three-level hierarchical tree decoder was utilized to first decode the large/small distinction (utilizing amplitude and spectral features), then the living/non-living object category (utilizing 200-700ms amplitude features), and finally the individual word (utilizing 250-500ms amplitude features). A) Average accuracies at each branch of the tree are shown with corresponding colors. B) Accuracies at each level of the decoder are shown on a per subject basis with dotted lines indicating chance accuracy. C) Cumulative accuracies at each level decrease as errors propagate through levels of the tree. D) Performance of the hierarchical tree is a significant improvement (Wilcoxon sign-rank, $p<0.05$) over training a single multi-class SVM to discriminate between all 10 words.

4. DISCUSSION

Understanding not only the spatial, but also the temporal representation of semantic categories and individual words requires analysis techniques robust to the high dimensionality of multichannel EEG and MEG data. In this chapter, we have demonstrated that a machine-learning technique, such as SVMs, can detect distributed differences in neural activity and robustly extract language-related information from electrophysiological recordings. These

representations are supramodal and are relatively consistent between individuals. Utilization of a scalable hierarchical classification construct allows us to decode various word properties sequentially and further improves decoding performance.

Previous imaging studies investigating the neural basis of living/non-living object representations have yielded variable results. The most consistent findings have been activation of the posterior middle temporal gyrus to tools and man-made objects and activation of inferior temporo-occipital cortex to animals (Moore and Price, 1999, Mummery et al., 1998, Mummery et al., 1996, Martin et al., 1996, Perani et al., 1995, Chao et al., 1999, Perani et al., 1999, Damasio et al., 1996). Despite this, inconsistencies exist in the literature. While only a few studies report entirely conflicting results, many of the brain areas identified as showing differential activation to living or non-living objects have only been reported in a single study. Moreover, one other MEG study failed to find any statistically significant differences between the perception of natural and man-made objects (Low et al., 2003). This may be due, in part, to experimental design, but may also be due to the statistical mapping analysis used to analyze these neuroimaging data. These techniques must correct for multiple comparisons and thus are most sensitive to brain areas which demonstrate large differences in activation between conditions. Often, multiple comparison corrections are based on spatial clustering, thus biasing the results toward contiguous arrays of activated voxels. The decoding of single channel data shown here demonstrates that the information from single sensors in insufficient to provide the same accuracy obtained when utilizing all sensors together, but that a large number of sensors contribute small amounts of information. This is consistent with the idea of a distributed representation of semantic knowledge, and may partly explain why statistical techniques that are most sensitive to large, concentrated effects may yield variable results if the distribution of information is instead widespread and subtle.

The SVM weights from our recordings are generally consistent with these previous imaging results, however, several important differences exist. The MEG data suggest that the bilateral anterior temporal, bilateral inferior frontal, and left parietal regions contribute to non-living object category representation from 400-500ms. Frontal activity sensitive to object category has been mainly observed the left hemisphere. There is growing evidence that the anterior temporal lobes play a crucial part in the representation of semantic knowledge, and the large SVM weights observed in this chapter are consistent with this hypothesis (Patterson et al., 2007, Lambon Ralph et al., 2009, Lambon Ralph et al., 2010). I will further explore the role of the anteroventral temporal lobe in semantic representations in Chapter III.

Also consistent with previous results, left inferior temporal-occipital SVM weights specific for living objects are apparent from 400-500ms, but early 200ms non-living object-specific weights are also present in the same area. Activation to inanimate objects has not previously been seen in left inferior temporal-occipital cortex via functional neuroimaging. This finding suggests that a single brain area may respond to living and non-living categories at different latencies, but more focal intracranial recordings may be necessary to further substantiate this hypothesis and will be explored in Chapter III.

Utilizing a recording modality with sufficient time resolution, and an analysis technique designed to handle high-dimensional data, allows for the discrimination of such time-separated effects. In contrast, the temporal blurring in fMRI and PET may only allow detection of the larger or more prolonged effect. This may explain discrepancies in previous imaging results; it is possible that the particular cognitive demands of each experimental task may elicit varying latencies of activity that manifest themselves differently in low time-resolution neuroimaging data.

Furthermore, while the N400 event related potential (ERP) is known to be modulated by various semantic effects (Kutas and Hillyard, 1984, Kutas and Hillyard, 1980, Holcomb and Neville, 1990), our results suggest that earlier components (possibly as early as 200-300ms) may also contribute to the encoding of object category. This is especially pronounced in left inferior temporal-occipital sensors at 200ms when classifying living versus non-living objects in both visual and auditory modalities.

The results presented here also suggest a potential structure to the underlying representation of individual words. Because extracranial electrodes record the activity of large networks of concurrently active neurons, it is possible that the word-specific responses observed in our data are the superposition of many specific neural responses to lexico-semantic features of each word, as others have suggested (Pulvermuller, 2005, Tyler and Moss, 2001, Tyler et al., 2000, Caramazza et al., 1990). For example, the neural response to the word "banjo" may be comprised of the sum of the specific activations related to how a banjo sounds, the visual characteristics of a banjo, the fact that "banjo" is a noun, and all the associative elements specific to the individual. One might expect that concepts with large amounts of overlapping characteristics would incorporate similar neural networks, and thus would have similar macro-scale representations. The confusion matrix analysis supports this idea by indicating that, while the classifier was able to decode individual words, fewer errors were made between living and non-living objects than within either of these semantic categories. This supports the intuitive notion that the representation of semantically related words may be more similar than the representation of words which are distant in semantic space. Under this hypothesis, it is not surprising that the novel trials separate nicely into living and non-living object categories.

The results of this chapter also demonstrate the ability to extract semantic category and individual word representations from spatiotemporal features generated from noninvasive neurophysiology. The successful inter-modal classification shows that our machine-learning models are able to extract semantic information that is not specific to a single sensory modality. Not surprisingly, the cross-modality decode performance was lower than single modality performance; this may be partially due to differences in sensor placement and cognitive state between performance of the SV and SA tasks, often separated by months.

Despite the variability in recording conditions, inter-subject classification also performed significantly above chance levels, suggesting that the representations of various objects, concepts, and semantic categories may be fairly consistent across individuals. However, it is important to note that the generalized decoder performs far worse than the subject-specific decoders. Variable electrode placement, variable cortical language-related representations, or both may account for this. Potential intersubject variability often decreases the sensitivity of traditional statistical mapping techniques used in imaging studies. Subject-specific decoding analyses, like the one presented in this chapter, overcome this by training and testing on a subjects own data.

A few other studies have attempted to decode word processing information from electrophysiology (Suppes et al., 1997, Suppes and Han, 2000, Suppes et al., 1999). These studies utilize specifically chosen single channels of EEG. The minimum-squared-error classifiers they used were therefore appropriate for decoding these low-dimensional data. In our case, however, feature vectors capture the entire spatiotemporal dynamics of each trial, and thus machine-learning techniques which are robust to high-dimensional data were necessary. We achieved higher average accuracies after averaging 5 trials than these previous

studies have reported after averaging 10 trials (Suppes et al., 1997). In another study, Gonzalez Andino et al. (2007) also utilized SVMs to decode multichannel EEG recordings related to word and image processing. While the reported accuracies are impressive, the authors perform discrimination between distinct classes of stimuli (written words, pseudowords, line drawings, and scrambled images), rather than the more difficult task of decoding conceptual categories within a single stimulus modality.

While language information was extracted from both EEG and MEG recordings, MEG-based features yielded significantly higher accuracies. This differential accuracy of MEG versus EEG may be simply due to increased numbers of sensors in the MEG modality. However, despite the lower performance of the EEG features, combining EEG and MEG features improved performance over either recording modality alone, indicating that the information provided by EEG and MEG is not completely redundant. This suggests that neither recording modality is strictly superior to the other, and that EEG and MEG each provide unique information regarding neural processes. This notion is supported by widespread evidence that MEG and EEG are sensitive to different neurophysiological processes (Cuffin and Cohen, 1979, Wolters et al., 2006, Cohen and Cuffin, 1983, Huang et al., 2007, Dehghani et al., 2010).

While we have demonstrated that support vector machines are able to extract distributed language information from EEG and MEG recordings, not all multivariate classification techniques are equally successful. A naïve Bayes classifier performs significantly poorer than the SVM, suggesting that analysis of EEG and MEG data requires algorithms which are robust against overfitting, and can handle high-dimensional data.

The use of this decoding analysis not only provides insight into the nature of distributed language processing, but also has implications for the development of a language-based

neuroprosthesis. Machine learning algorithms, such as SVMs, can be trained on a patient's own data, making individual variability inconsequential. Furthermore, SVMs are robust to high-dimensional data, allowing for successful decoding broadly distributed semantic representations.

It is important to note, however, that the results presented here are extremely preliminary with regard to the development of a practical communication prosthesis. Various practical barriers must be overcome before a language prosthesis is viable. The tasks used here are language comprehension tasks while a language prosthesis involves language production. However, in most models of language processing, the same underlying semantic representations of each of word are activated in both production and comprehension (Martin, 2003, Dell and O'Seaghdha, 1992, Patterson et al., 2007, Indefrey and Levelt, 2004). We have demonstrated that the representations we are decoding are supramodal, suggesting that semantic content is a major source of information in these recordings. This semantic representation is the desired decoding target for a language prosthetic device, so utilizing these language comprehension tasks as an initial pass in decoding analysis is not unreasonable. Furthermore, others have reported an ability to decode the motor commands associated with articulation, and we believe incorporating semantic information, as seen in this chapter, may greatly benefit such efforts (Kellis et al., 2010, Guenther et al., 2009). This idea is explored briefly in Chapter V.

An algorithm which narrows the search space of possible words by first determining various word properties (grammatical class, semantic category, visual attributes, etc.) before decoding individual concepts may require much less training and provide an extensible classification algorithm. We have shown that this is possible with a hierarchical decoding

framework, and that performance improves as a result. The hierarchical framework presented here would allow for the decoding of a large library of concepts given the appropriate features to sequentially divide the search space. For example, concrete nouns and verbs produce different patterns of synchrony in EEG recordings (Weiss and Mueller, 2003), making coherence features a logical choice for discriminating this grammatical distinction. Given an adequate number of such distinctions, more realistically sized vocabularies may be utilized. The inclusion of a probabilistic syntactic/semantic language model, such as those used in automatic speech recognition (Baker, 1975), may further assist in narrowing the search-space and facilitate improved communication when moving beyond the decoding of single words. While, in general, imposing hard constraints on a learning algorithm reduces its flexibility and potentially prevents it from examining all pieces of data simultaneously, in this case, the hierarchical framework presented here can help to deal with limited amounts of data and provide extensibility. A variety of other methods may be used to combine multiple classifier models, including using meta-classifiers that take as input the discriminant scores of a "committee" of other classifiers to make a final decision.

The decoding analyses used in this chapter allow for the study of distributed, but potentially subtle, representations of semantic information within the human cortex. These multivariate techniques offer advantages over traditional univariate statistical mapping analyses. We have shown that high-dimensional machine-learning techniques, in conjunction with EEG and MEG recordings, provide insight into both spatial and temporal aspects of language processing. Furthermore, the ability to decode living/non-living category or individual words between subjects and stimulus modality suggests that these representations are consistent and supramodal. We have also shown that utilizing word property information

in an informed manner to decode individual words provides a potential framework for decoding larger libraries of words or concepts.

Given the recent evidence suggesting the importance of the anterior temporal lobes (ATL) in coordinating this distributed semantic information (Patterson et al., 2007, Lambon Ralph et al., 2010), consistent with the large anterior temporal SVM weights observed here, examining spatiotemporal aspects of processing during a semantic task in this area may be insightful. While the broad spatial coverage of EEG and MEG is well-suited for the study of distributed representations, it is not ideal for the study of the anterior inferior temporal lobes at finer spatial scales. In the next chapter, in order to investigate the semantic processing that occurs in the ATL, we shall move inside the skull and record from intracranial EEG electrodes.

III. EARLY SEMANTIC INFORMATION IN HUMAN ANTEROVENTRAL TEMPORAL LOBE

1. INTRODUCTION

While the previous chapter explored the distributed representations of semantic knowledge, in this chapter, we focus on one particular area that may be crucial for the coordination of this widespread information. EEG and MEG provide excellent coverage of most of the cortical surface, however, they are unable to reliably record activity from the inferior and medial temporal surfaces. Inferior temporal lobe is known to be a crucial part of the ventral visual stream for object recognition, and a portion of it may be equally important in the amodal representation of objects (Binney et al., 2010, Grill-Spector et al., 2004, Halgren et al., 1999, Haxby et al., 2001a, Mishkin et al., 1983). In this chapter, intracranial recordings are used to record from anterior inferior and medial temporal cortices to examine the differences in representation of two important semantic categories.

Evidence for selective effects of lesions on the ability to comprehend and produce words associated with living animals versus manmade objects has been reported since at least 1946 (Nielsen, 1946, Warrington and Shallice, 1984, Warrington and McCarthy, 1983). Deficits in production and comprehension of animal-related concepts are associated with lesions of the left inferior temporal and ventral occipital cortex, while deficits in naming and comprehending tools and manmade objects are associated with lesions of frontal premotor and posterior middle temporal gyri (McCarthy, 1995, Tranel et al., 1997, Mahon and Caramazza, 2009). Building on these findings, fMRI and PET studies have shown increased activity in lateral posterior fusiform gyrus and ventrolateral occipital cortex for animal versus tool stimuli (Chao et al., 1999, Martin et al., 1996, Hauk et al., 2008, Devlin et al., 2005), while manmade objects

evoke increased activation of middle temporal gyrus (Perani et al., 1999, Hauk et al., 2008) and medial fusiform gyrus (Chao et al., 1999, Devlin et al., 2005).

The selective hemodynamic responses to pictures depicting objects versus animals in the ventral occipitotemporal cortex have been interpreted as examples of high-level visuo-perceptual areas specialized for various categories of objects including faces and buildings (Martin, 2007, Kanwisher et al., 2001). Some neuroimaging studies have found that this selective response to animals versus objects extends to the words that refer to them (Chao et al., 1999, Devlin et al., 2005), others, however, have not (Phillips et al., 2002, Price and Devlin, 2003, Mummery et al., 1998). Since the tasks where words are effective are thought to invoke elaborative processing, and the areas involved are also activated by visual imagery (Ishai et al., 1999), the hypothesis has been advanced that this area is involved in structural, rather than semantic representations (Devlin et al., 2005). That is, in tasks which invoke extended processing of words, top-down projections from semantic areas are hypothesized to activate ventral occipitotemporal areas specialized for the perceptual processing of objects and animals (Devlin et al., 2005, Noppeney et al., 2006), whereas images activate this area in a bottom-up fashion (Mechelli et al., 2004, Mechelli et al., 2003, Noppeney et al., 2006). Devlin notes that the posterior occipitotemporal lesions producing category-selective visual agnosia often spare general semantic knowledge concerning the same categories, implying that the region is higher order visual-perceptual rather than semantic per se (Arguin, 1996, Etcoff et al., 1991, Humphreys et al., 1997)

This hypothesis predicts that words would evoke a relatively early and selective response in the semantic area in order to evoke the later, selective, feedback activation. Hemodynamic methods lack the temporal resolution to determine the latency of category-

specific activity. Here, we utilize intracranial EEG (iEEG) to study category-selective responses in anteroventral temporal lobe (avTL) to words referring to animals and manmade objects. Utilizing written and spoken words, as well as multiple tasks, allows for the exploration of supramodal, task-independent semantic representations. Microelectrode and macroelectrode arrays provide the spatiotemporal resolution for the detection of early, potentially first-pass, category-specific activation in ventral temporal lobe.

2. METHODS

2.1. PARTICIPANTS

Nine patients, five female and four male, at the Massachusetts General Hospital or Beth Israel Deaconess Medical Center with medically intractable epilepsy participated in this research while undergoing clinical evaluation using intracranial electrodes. Patients were between ages of 17 and 65 and were all right handed. As discussed below, patients were implanted with a variable number of depth electrodes as determined by a clinical team caring for the patients. Patients were enrolled in this study under the auspices of local IRB oversight in accord with the declaration of Helsinki. See Table III-1 for detailed subject information.

TABLE III-1: PATIENT INFORMATION

Patient	Electrodes	Location	x	y	z	Handed-ness	Language Dominance	Sex	Age	Tasks	Etiology	Seizure Onset Zone	Resected Area	Clinical Outcome
D1	Depths	R-HC/PHc	20	-16	-25	R	Left based on handedness	F	55	SV	Mesial temporal sclerosis	Bilateral mesial temporal lobes	None	Medical Management
D2	Depths	L-HC/ColS R-OTS/ITS R-HC/ColS	-37 45 39	-37 -24 -35	-13 -14 -9	R	Left based on handedness	F	65	SV	Unknown	Right subfrontal	None	Medical Management
D3	Depths	R-OTS R-OTS	41 38	-19 -33	-26 -13	R	Left based on handedness	M	23	SV	Unknown	Left frontal	None	Medical Management
D4	Depths	L-OTS R-ColS	-45 38	-33 -33	-14 -16	R	Left based on handedness and Wada	F	42	SV	Unknown	Left fronto-temporal	None	Medical Management
D5	Depths	L-OTS L-OTS L-HC/ColS R-OTS	-41 -42 -26 41	-11 -33 -11 -11	-28 -14 -28 -29	R	Left based on handedness and Wada	M	29	SV, SA	Neurons in frontal white matter	Right fronto-temporal	R. ant. temp. lobectomy and R. front. corticectomy	Recurrent seizures
D6	Depths	L-ColS R-HC/TOS	-38 37	-16 -13	-16 -24	R	Left based on handedness and MEG	F	27	SV, SA	Cortical Dysplasia	Left frontal	Left frontal corticectomy	Seizure free at 4 months
L1	Laminar microelectrode	R-IT				R	Left based on handedness	M	30	SV, WM	Resection not revealing	Left frontal	Left frontal resection	Rare seizures
L2	Laminar microelectrode	R-PR				R	Left based on handedness	F	35	SV, WM	Unknown	Right orbitofrontal	None	Medical management
L3	Laminar microelectrode	L-ER				R	Left based on handedness	M	35	SV, SA, WM, DI	RHC sclerosis	Right amygdala	Right temporal lobectomy	Seizure free

2.2. INTRACRANIAL ELECTRODES AND RECORDING

Intracranial EEG (iEEG) recordings were obtained from up to 80 channels of clinical macroelectrode arrays. Six or eight contact depth electrodes, manufactured by Adtech Medical (Racine, WI), were utilized to record from both medial and lateral cortical areas of the frontal and temporal lobes. Contacts were platinum cylinders, 1.1mm in diameter and 2.3mm in length, with 5mm between the center of adjacent contacts. The decision to implant electrodes and the type, number and spatial configuration of electrode placement was determined entirely on clinical grounds. iEEG was continuously recorded at 500Hz with band-pass filtering from 0.1 to 200Hz. These macroelectrode depth recordings were obtained from six of the nine patients.

Intracranial macroelectrodes were localized by using a volumetric image coregistration procedure (Dykstra et al., 2012). Using Freesurfer scripts (http://surfer.nmr.mgh.harvard.edu), the pre-operative T1-weighted MRI (showing the brain anatomy) is aligned with a post-operative CT (showing electrode locations), and both are transformed into Talairach coordinates. Electrode coordinates were manually determined from the CT and also placed into Talairach space. To visualize electrode locations, coordinates were plotted on the average Freesurfer pial surface (fs-average) and individual coronal MRI slices were obtained for each contact.

The remaining three patients were implanted with linear arrays of microelectrodes capable of recording local field potentials (LFPs) and multi-unit activity (MUA) across the cortical layers. These arrays were 3.5mm in length with 24 platinum-iridium contacts (40μm diameter) spaced 150μm apart. Recordings from these laminar electrodes were obtained in dual bands: 2kHz sampling rate for field potentials, and 20kHz for unit activity. The amplifier

utilized a bipolar electrode configuration to minimize noise. For details of construction and use of these arrays, refer to Ulbert et al. (2001), Cash et al. (2009), Csercsa et al. (2010) and Keller et al. (2010). In these three patients, post-operative T1 and T2 weighted MRIs were obtained with the electrodes in place. Direct visualization localized the microelectrodes to the inferotemporal cortex (IT) in patient L1, perirhinal cortex (PR) in patient L2, and entorhinal cortex (ER) in patient L3 (Table 1). Localization in the MRI, informed by the known laminar cytoarchitecture of the respective cortical areas, and confirmed and refined by determination of background activity (e.g., white matter and CSF have different amplitude local field potentials as compared to gray matter) permitted the individual contacts of each laminar electrode to be assigned to putative cortical layers (Ulbert et al., 2004a, Ulbert et al., 2004b, Halgren et al., 2006, Fabo et al., 2008).

2.3. ANALYSIS

Averaged local field potentials (LFPs) were computed for all macroelectrode recordings. Continuous signals from each iEEG channel were initially low-pass filtered at 30Hz and subsequently epoched from one second before to two seconds after stimulus onset. Trials containing large artifacts were rejected using a predefined amplitude threshold, and trials containing epileptic discharges were rejected manually. After alignment to stimulus onset, waveforms from all channels were baseline corrected using a 500ms pre-stimulus period. These preprocessing steps were performed within MATLAB, using the EEGLAB 6.03b toolbox (Delorme and Makeig, 2004).

Gamma-band responses were also computed for macroelectrode recordings. Power was first computed from 30 to 100Hz in 2Hz increments using a Morlet wavelet time-frequency analysis. The number of wavelet cycles was increased linearly from 3.6 to 12 as frequencies

ranged from 30 to 100Hz, providing a constant temporal and frequency resolution across the entire band. The resulting temporal resolution (σ_t) was 30ms with a corresponding frequency resolution (σ_f) of 8Hz. Spectral power at each frequency was normalized to the power in a 500ms pre-stimulus baseline before averaging across the entire band to generate a single gamma-band event-related spectral perturbation waveform.

From the laminar microelectrodes, population current source density (CSD) and multi-unit activity (MUA) were estimated. The CSD estimates the transmembrane current in each cortical layer, while MUA estimates changes in firing rate of the same population of neurons. CSD was computed as the second spatial derivative of field potentials after applying a 5-point Hamming filter (Ulbert et al., 2001). Differential transmembrane current sources between conditions were displayed by plotting the subtraction of the mean CSD for objects from the mean CSD for animals. When comparing between more than two conditions, the F-statistic from a one-way ANOVA of the CSD computed across individual trials was plotted as a measure of the difference between conditions. MUA was computed by first filtering the 20kHz signal from each channel between 500 to 3000Hz and subsequently rectifying the signal. This rectified signal was then low-pass filtered at 30Hz.

To test the statistical significance of response differences between animal and object categories, a cluster-based non-parametric Monte-Carlo hypothesis test was used on LFP and MUA waveforms, gamma-band power, and CSD plots (Maris and Oostenveld, 2007). This corrects for multiple-comparisons while preserving sensitivity in the time-domain. All reported temporal regions of significant differences within averaged LFP, gamma waveforms, CSD or MUA plots are at a $p<0.05$ level.

To quantify and study the information content contained in the microelectrode recordings, support vector machine classifiers were trained and tested using features computed from CSD and MUA. A two-class SVM was used with a radial-basis function kernel with parameters C=10 and γ=0.01 determined by cross-validation within the training set. To evaluate the accuracy of the SVM with a given set of features, 10-fold cross validation was performed. Extracted features included averaged CSD and MUA for each channel within a specified window. Depending on the analysis, various sizes and numbers of windows were used. For example, for quantifying information in time, a single sliding window, or a set of cumulative windows, were used to train and test the SVM at a number of latencies.

To study the firing of single-units in the microelectrode recordings, signals were high-pass filtered at 250Hz, and a threshold was automatically chosen as 4 times the standard deviation of the signal. 1600ms of action potential waveforms that crossed the threshold were extracted, and manually clustered in the 2D space defined by the first two principal components. Filtering, thresholding, and spike sorting was performed using Offline Sorter (Plexon, Dallas TX).

2.4. LANGUAGE TASKS

All participants performed a language task involving written words (SV) while two participants also performed an auditory-word version (SA) of the same task. These tasks presented words referring to animals or objects. To require access of semantic information of these words, the subject was asked to press a button if the object being referred to was larger than a foot in any dimension. These tasks are described in detail in Section II.2.2.

A word memory task (WM) was also performed on the three subjects implanted with microelectrode arrays. Subjects were first asked to remember a list of 10 words, each presented 3 times. During the experiment, words were visually presented for 300ms with a stimulus onset asynchrony of 2000ms. Subjects were asked to press a button whenever any word from the initial list was visually displayed. The target words were shown 12 times each (for a total of 120 trials) while 120 novel words were displayed only once over the course of the experiment. Words were either animals, manmade objects, or abstract nouns (e.g. "respect", "honor").

Patient D5 performed a picture naming task (PN) in which he was shown pictures corresponding to concrete objects in one of eight categories: dogs, cats, roses, trees, balls, bats, cars, and boats. These images were displayed for 300ms on screen and subsequently extinguished. After 450-550ms, a go cue would be displayed on the screen and the subject would be required to say the category of the picture. All images were only shown once and were normalized for intensity and contrast.

Finally, patient L2 also performed an abstractness judgment task (DI) in which the subject was asked to view visually presented words and respond to any words that were "abstract" (e.g. love, honesty) rather than "concrete" (e.g. cat, house). A total of 480 novel words were presented, with no repetition, and words referred to animals, objects, or abstract nouns.

3. RESULTS

3.1. AVERAGED LFP DIFFERENCES BETWEEN ANIMALS AND MANMADE OBJECTS

In general, averaged LFP waveforms in anteroventral temporal lobe showed large deflections at approximately 400-500ms (Figure III-1). This is likely an intracranial manifestation of the well-studied scalp N400 potential (Kutas and Federmeier, 2000, Marinkovic, 2004).

Robust animal/object specific activity was observed in bilateral ventral and medial temporal areas in the averaged LFPs of the six patients with macroelectrode recordings (Figure III-1, solid-lined plots). Specifically, electrodes in or near collateral and occipito-temporal sulci, both anteriorly and posteriorly (relative to the span of recorded sites), demonstrated category-specific activity. Two electrodes near right hippocampus and parahippocampal gyrus also showed category-specific activity. Significant differences between the two semantic categories were observed as early as 200ms and as late as 1500ms. While category-specific differences were apparent at the 400-500ms peak in seven of the thirteen electrodes, six electrodes demonstrated differences within the slow return to baseline beyond 500ms. In all but two cases, the right hemisphere electrodes in patient D3, the response to animals yielded more negative LFP waveforms than the response to manmade objects.

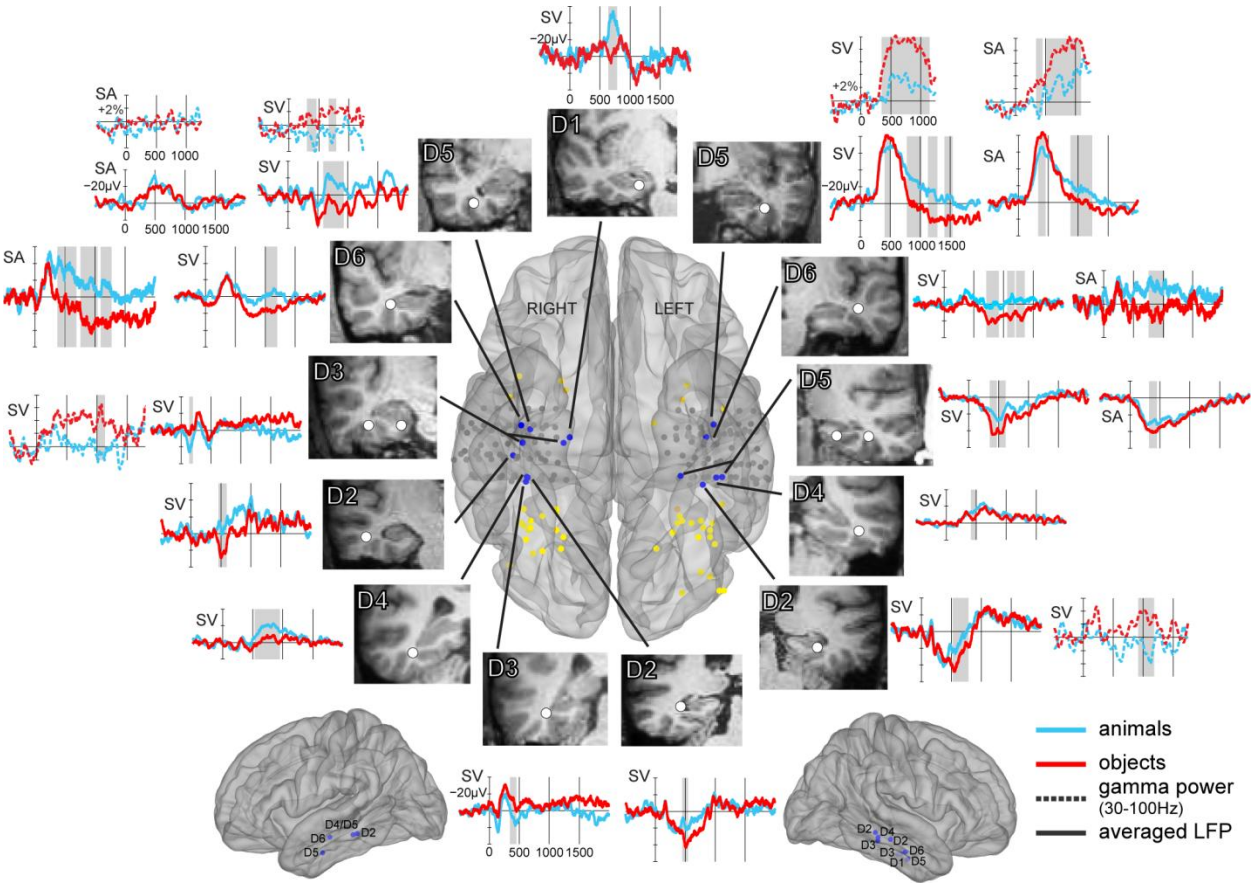

FIGURE III-1: VENTROTEMPORAL CATEGORY SPECIFICITY IN AVERAGED LOCAL FIELD POTENTIALS

(center) Depth electrode coordinates from all patients in Talairach space plotted on the Freesurfer average surface. Blue electrodes indicate temporal recording sites demonstrating significant averaged LFP differences, gray electrodes indicate temporal recording sites without significant LFP differences, and yellow dots indicate Talairach coordinates of either the center or maximally significant voxel for category-specific fMRI or PET responses as reported in previous literature. Coronal MRI slices of the temporal lobe are shown for each significant electrode location. (sides) Averaged LFP waveforms (solid lines) or gamma power (dashed lines) for animals (blue) versus objects (red). Electrodes in occipito-temporal sulcus (OTS), collateral sulcus (ColS), and hippocampus/parahippocampal gyrus (HC/PH) demonstrate category specificity. Differences are seen largely starting at 400ms and in some cases, remain until 1500ms after stimulus onset. In four subjects, gamma-band power (30-100Hz) was differentially modulated by animals and objects. Latencies of significant differences are seen as early as 300ms and as late as 1200ms.

In the two subjects who also performed the auditory version (SA) of the size judgment task, D5 and D6, the electrodes which demonstrated written word category specificity also exhibited category-specific responses to spoken words. The averaged LFP waveforms in both

task modalities exhibit similar morphology, and category-specific differences occur at similar latencies. This suggests that these ventral temporal regions are supramodal with respect to the encoding of semantic category.

Patient D5 also performed a picture naming task in which the subject was shown a series of images that corresponded to one of eight categories of objects: cats, dogs, roses, trees, balls, bats, cars, and boats. High gamma power (HGP) in the left anterior occipito-temporal sulcus electrode demonstrated larger increases in response to images of manmade objects (balls, bats, cars and boats) and smaller changes in response to animals and plants (dogs, cats, roses and trees) (Figure III-2). This is consistent with the category-specific differences observed in response to words at this electrode, and demonstrates that these category-effects are present for non-lexical stimuli that activate the same semantic information. This category difference is less apparent in the LFP, but interestingly in both LFP and high gamma power, statistically significant differences occur at approximately the same latencies, although significant differences in HGP begin slightly later at 450-500ms as opposed to 350-400ms for the lexical stimuli.

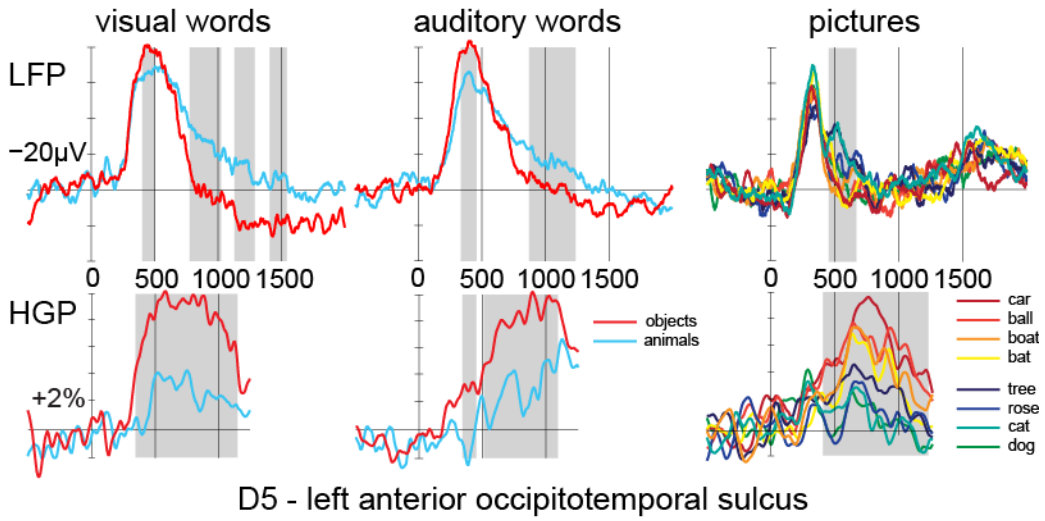

FIGURE III-2: RESPONSES TO PICTURES ALSO DEMONSTRATE CATEGORY-SPECIFICITY

Category-specific differences are apparent in the visual and auditory word stimuli as seen previously, but are also apparent in picture stimuli in the high gamma power (HGP). These differences are less clear in the LFP traces for picture stimuli, but the latency of category-specific differences are similar for both words and pictures.

The location of the electrodes demonstrating category-specificity found here are further anterior than the posterior ventral temporal locations reported in hemodynamic studies of activation to pictures or words representing objects or animals (Thompson-Schill et al., 1999, Chao et al., 1999, Devlin et al., 2005, Perani et al., 1999, Noppeney et al., 2006, Price et al., 2003a, Mechelli et al., 2006, Whatmough et al., 2002) (Table III-2). While the Talairach coordinates of fusiform-specific category-selective activity reported in those studies ranged from y=-33 to -83 (mean = -58) in the anterior-posterior axis, the coordinates of the involved electrodes in this chapter ranged from y=-11 to -37 (mean = -24). In four other studies, PET or fMRI category-specific activity was observed in response to pictures at the temporal poles (Mummery et al., 1996, Damasio et al., 1996, Devlin et al., 2002, Moore and Price, 1999). This activity is further anterior than the recording sites reported here. Category-specific activity in this portion of the ventral temporal lobe has not previously been reported by neuroimaging.

Because these category-specific findings are generally consistent across subjects despite the varying epilepsy etiologies and seizure foci (Table III-1), it is likely that the semantic processing observed in the avTL is representative of the normal function of this area.

TABLE III-2: TALAIRACH COORDINATES OF CATEGORY-SPECIFIC RESPONSES IN PREVIOUS NEUROIMAGING STUDIES

Study	Contrast	Modality	x	y	z
Chao et al. (1999)	animals>tool (viewing)	fMRI	38	-56	-12
	animals>tool (viewing)		-40	-59	-10
	animals>tool (matching)		41	-53	-20
	animals>tool (matching)		-35	-59	-20
	animals>tool (naming)		37	-52	-20
	animals>tool (naming)		-37	-55	-20
	animals>tool (reading)		37	-55	-21
	animals>tool (reading)		-40	-56	-21
	tool>animals (viewing)		26	-48	-9
	tool>animals (viewing)		-26	-47	-5
	tool>animals (matching)		32	-65	-19
	tool>animals (matching)		-26	-53	-17
	tool>animals (naming)		26	-47	-16
	tool>animals (naming)		-27	-50	-15
	tool>animals (reading)		23	-59	-11
	tool>animals (reading)		-32	-53	-17
Thompson-Schill et al. (1999)	living>nonliving (visual questions)	fMRI	-41	-53	-11
	living>nonliving (nonvisual questions)		-45	-45	-8
	nonliving>living (visual questions)		-41	-53	-11
Perani et al. (1999)	living>nonliving (discrimination)	PET	-44	-82	-32
	living>nonliving (discrimination)		-46	-82	-20
	living>nonliving (matching)		-28	-83	-16
	living>nonliving (matching)		-36	-74	-12
Whatmough et al. (2002)	tools>animals	fMRI	24	-64	-9
	tools>animals		-17	-64	-12
	animals>tools		40	-47	-17
	animals>tools		46	-71	-5
	animals>tools		-40	-76	2
Price et al. (2003)	natural>manmade	fMRI	-42	-62	-20
	natural>manmade		40	-54	-14
Devlin et al. (2005)	natural>manmade	fMRI	-36	-52	-18
	manmade>natural		26	-56	-10
	natural>manmade		36	-66	-14
Noppeney et al. (2006)	tools>animals	fMRI	-24	-57	-15
	tools>animals		-33	-33	-24
	animals>tools		39	-60	-21
Mechelli et al. (2006)	artifacts>animal (relevance)	fMRI	-28	-52	-14
	artifacts>animal (relevance)		32	-50	-16
Mummery et al. (1996)	natural>manmade	PET	-16	-10	-16
	natural>manmade		22	4	-12
Moore et al. (1999)	natural>manmade	fMRI	-28	0	-14
	natural>manmade		42	10	-18
Devlin et al. (2002)	living>manmade	PET	24	8	-24
	living>manmade		-30	6	-18
Damasio et al. (1996)	faces>animal+tools	PET	46	1	-27
	faces>animal+tools		-46	-4	-28

3.2. GAMMA-BAND SELECTIVITY

In three of the subjects, category-specificity was found in gamma-band power (30-100Hz) in medial and inferior temporal electrodes (Figure III-1, dash-lined plots). In these subjects, gamma-band power increases at approximately the same time as the major deflection of the averaged LFP. However, in several cases, these responses continue beyond the 400-500ms peak in the field potentials, demonstrating that increased gamma power may continue even after the field potential returns to baseline. Time-frequency plots of these channels indicate that the high-frequency activity observed here is a result of increases in frequencies between 30 to approximately 120Hz. The most pronounced example of category-specificity is seen in subject D5. In this subject, gamma-band category differences are clearly visible in both visual and auditory modalities of the size judgment task in the left anterior occipito-temporal sulcus electrode. While significantly more gamma-power is visible in response to object trials, gamma-power increases for both semantic categories. Significant differences begin at 300ms and last until 1200ms. These latencies begin slightly earlier than the corresponding LFP differences observed in the same electrode.

Gamma-band and LFP specificity, although spatially correlated, were not simultaneously present for every electrode. In patients D1, D4 and D6, none of the channels showing LFP category specificity showed differential gamma-band activity. In patients D2 and D3, one of the three electrodes that showed LFP specificity also showed gamma-band specificity. In patient D5, one electrode showed only LFP specificity, one showed only gamma-band specificity, and two showed specificity in both types of activity. These data demonstrate that while LFP and gamma-band activity often occur together, they can also occur independently. In all electrodes showing gamma band differences except for one, LFP differences were also seen, suggesting that gamma activity tends to be more focal.

3.3. MULTIUNIT ACTIVITY AND CURRENT SOURCE DENSITY

In the three subjects with laminar microelectrode arrays, CSD plots illustrate robust task-related responses in IT, PR, and ER (Figure III-3). Category-specific differences are observed in the size judgment (SZ), abstractness judgment (DI), and word memory (WM) tasks. This suggests that even in a task which does not require explicit access of visual-structural information (the word memory task), category-selective responses are still observed.

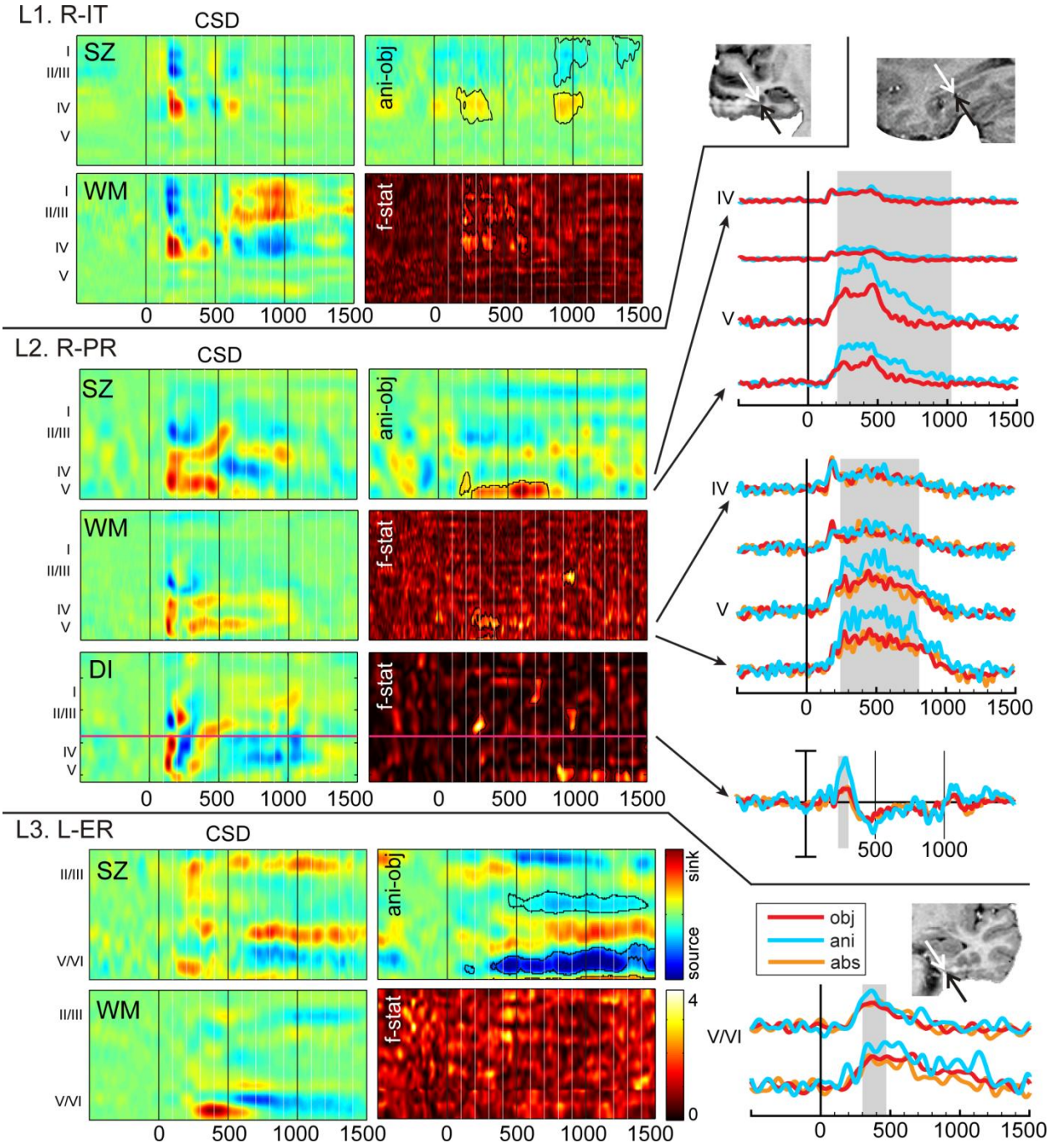

FIGURE III-3: LAMINAR MICROELECTRODE RECORDINGS DEMONSTRATE CATEGORY-SELECTIVE RESPONSES

Current source density (CSD) and multi-unit activity (MUA) show category-specific differences between animals and manmade objects for the three implanted patients. CSD was computed as the second spatial derivative of laminar recordings. In CSD plots, outlined regions indicated statistically significant differences between animals and objects for the SZ task, or animals, objects and abstract nouns in the WM or DI tasks ($p<0.05$). 'ani-obj' plots were generated by subtracting the mean CSD for objects from the mean CSD for animals. Plots of the F-statistic from a one-way ANOVA indicate differences between three conditions (object/animal/abstract) for the WM or DI

(Figure III-3 continued) tasks. In MUA waveform plots, shaded regions indicate time-points with statistically significant differences. (L1) The right IT electrode shows a layer IV sink beginning at 160ms that is modulated by semantic category in both SZ and WM tasks. (L2) In the right PR electrode, the first sink occurs around 100ms in layers IV/V in all three tasks. Category-specificity is seen in these same layers beginning as early as 150ms. Differential MUA responses are seen in deeper layers and demonstrate animal-specific increases in firing beginning as early as 200ms. (L3) In the left ER electrode, an initial layer V/VI sink is present beginning as early as 100ms in the SZ task and around 200ms in the WM task. Category-selectivity is present in deeper layers at 130ms and more superficial layers later. MUA responses for the WM task demonstrate animal-specific increases in firing.

In the inferotemporal electrode in patient L1, activation begins with a sink in putative layer IV with a concurrent source in layers II/III peaking at 160ms in both SZ and WM tasks. Category-specific differences are seen within this first layer IV sink starting at 150ms, and again at 900ms in both layer IV and upper layers. This difference can be characterized by a larger layer IV sink in response to animals.

In the right perirhinal cortex electrode in patient L2, an early sink is again present in putative layer IV beginning at 120ms and peaking at 150ms followed by a superficial layer II/III sink at around 500ms for the SZ task. Category differences are observed within this first activation in layers IV/V starting at 150ms, and are again characterized by a larger sink in response to animals. Early responses to the WM and DI tasks are very similar in terms of laminar distribution, latency, and category-specific difference.

Patient L2 also yielded reliable multiunit activity for SV and WM tasks. Robust increases in MUA are apparent in layers IV and V with clear differences between animals and objects. This increase in unit-firing implies that the early layer IV/V sink in the CSD is excitatory in nature. The MUA response to animals is significantly larger than the response to manmade objects with differences beginning around 200ms. Similarly, in the WM task, the MUA response to animals was significantly larger than the response to objects or abstract nouns, but no difference was found between these latter two categories ($p>0.05$). These

differences begin at 230ms. In the abstract judgment task, the CSD response to animals was again larger than either man-made objects or abstract concepts, with no difference between the two latter categories, mirroring the MUA differences in the SZ and WM tasks.

In the entorhinal cortex electrode in patient L3, activation begins with a sink in layer V/VI at 120ms followed by a sink in superficial layers II/III around 190ms. In the ER electrode, differences are seen starting at 130ms in deeper layers with additional differences appearing in more superficial layers around 450ms. These differences are quite prolonged, and last beyond 1500ms. Robust multiunit activity was also observed in layers V/VI for the WM task in this patient. As in the case of the perirhinal electrode, the activity in this entorhinal electrode increases over baseline, indicating an excitatory early sink. Again, the MUA shows a larger increase to animals but no differences between manmade objects and abstract objects.

In all cases, category-specific differences occur at the layer and latency of the first current sink in the CSD, suggesting that first-pass activation of these areas contains semantic information. Furthermore, the layer IV location of this initial current sink in IT and PR is consistent with the typical layer where feed-forward activation arrives. This also suggests that the main source of this information is from longer-distance cortico-cortical afferents rather than local interneurons.

3.4. DECODING OF SEMANTIC CATEGORIES

To quantify the semantic information present in the microelectrode recordings, we performed a decoding analysis using Support Vector Machines as described in Section I.3. Similar to the feature extraction in Section II.2.4, the mean CSD or MUA values in four windows

from 100-800ms were used as the features for the SVM classifier. Features from all channels were concatenated to form the final feature vector.

Data from all three microelectrode patients allowed for significant decoding performance of animal or object category. Accuracies approached 95% in patient L1, 85% in patient L2, and 70% in patient L3 when using only CSD features (Figure III-4). When using MUA features in patient L2, accuracies reached 94%. Also, as expected, with increasing numbers of averaged trials, the accuracy increased. Most of the gains in accuracy could be achieved after averaging 10-15 trials. Furthermore, by combining both MUA and CSD features in patient L2, decoding performance improved as seen with the combination of EEG and MEG in Section II.3.3.

By examining the SVM weights, we are able to determine the CSD/MUA features that the classifier relied most heavily upon. In the case of the perirhinal microelectrode (L2), deep MUA and CSD features provided the most information over a wide range of latencies. The entorhinal electrode in patient L3 demonstrated that both deep and superficial layers provided the most information in the time window from 250-400ms, although in the earlier time window from 100-250ms, the deeper layers provided the majority of the contribution to the classification. The IT electrode in patient L1 showed a scattered distribution of SVM weights, a predominance in deep layers early on followed by more superficial layers contributing to the classification later. In general, the significant weights for all three microelectrodes reflect the observed patterns of animal/object significance seen in Figure III-3.

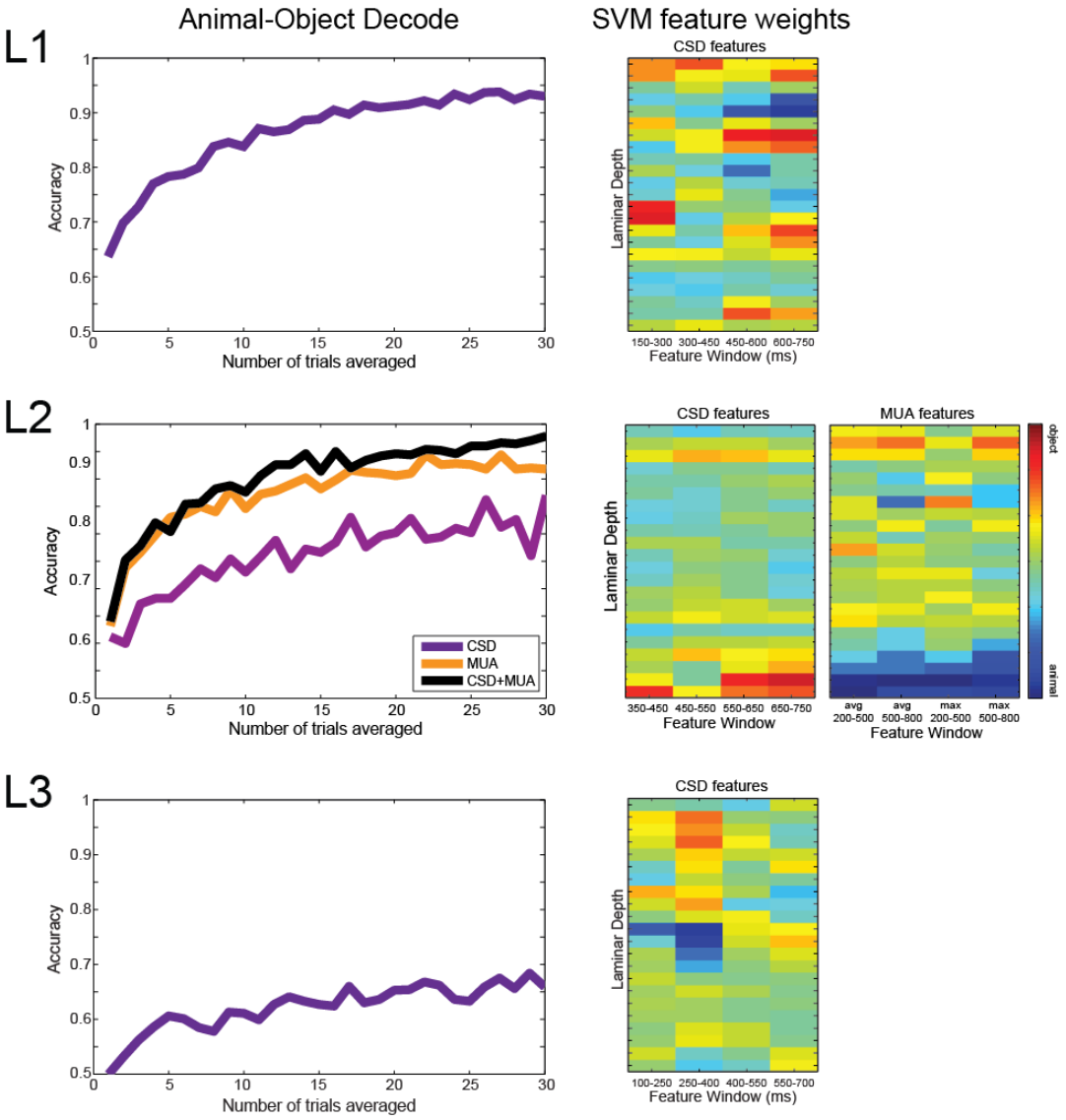

FIGURE III-4: DECODE PERFORMANCE USING FEATURES FROM LAMINAR MICROELECTRODE.

Left) The decode accuracy as a function of the number of trials averaged when attempting to discriminate between animal and object categories. Purple lines indicate performance when using CSD features, orange indicates the use of MUA features, and black indicates the use of both types of features. Right) The SVM feature weights for CSD and MUA features reflect early contributions to the final classifier model.

To quantify the time evolution of semantic information in the microelectrode recordings, we performed a decoding analysis of consecutive time windows (Figure III-5). In the first case, animal or object category was decoded using cumulative 25ms time windows

such that by the end, the time windows spanned from 0 to 1000ms. In the second analysis, a sliding 50ms window (overlapped by 40ms) quantified the information existing at particular time points. Both cumulative time windows (Figure III-5A) and sliding windows (Figure III-5B) demonstrate that significant semantic information is present as early as 150ms in all three microelectrode arrays. In the sliding window analysis, it is also clear that a cluster of information is present from approximately 150-350ms, and another cluster of information at 500-600ms. These plots also make it clear that the MUA features of the perirhinal electrode provide more semantic information than the CSD features.

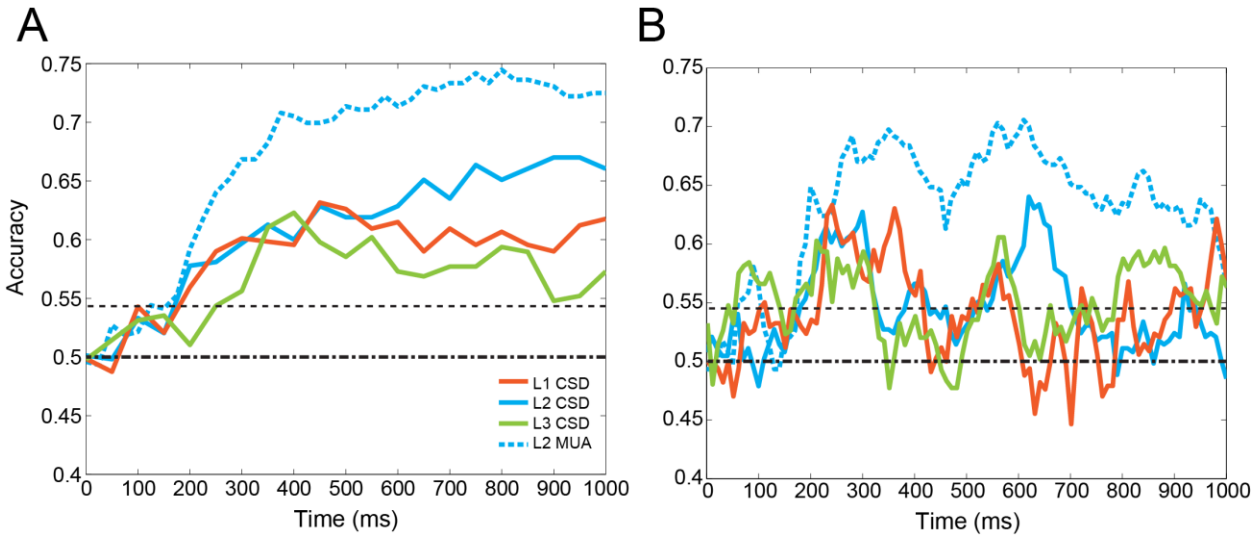

FIGURE III-5: DECODE PERFORMANCE OF CSD AND MUA FEATURES DEMONSTRATE EARLY SEMANTIC INFORMATION.

A) The decoding of animal or object category from CSD (solid lines) or MUA (dashed line) features computed in cumulative 25ms windows. The x-axis is the time of the last window used for the decoding (e.g. a time of 250ms indicates that 10 windows from 0ms to 250ms were used in the decoding). The dash-dotted line indicates chance accuracy (50%) while the dashed line indicates the significance threshold ($p<0.05$). B) The decoding of animal or object category using 50ms sliding window features. Therefore, the x-axis value indicates the beginning of the window used to obtain the corresponding decoding accuracy. Windows were shifted by 10ms (overlapped by 40ms).

3.5. SINGLE UNIT CATEGORY SELECTIVITY

Single-unit firing was identified in the perirhinal microelectrode recordings of patient L2 (Figure III-6). A total of eight distinct units were identified across the 24 channels. A raster plot of a representative unit is shown in Figure III-6A. In this case, firing decreased after stimulus onset. While mean firing rates were low (~0.1Hz), three of the eight units demonstrated statistically significant differences in firing for animal and object trials between 0 to 300ms (Wilcoxon rank-sum, $p<0.01$) (Figure III-6C). In all three cases, more spiking was observed in response to animals than objects, which is consistent with increased MUA response to animals in the same electrode.

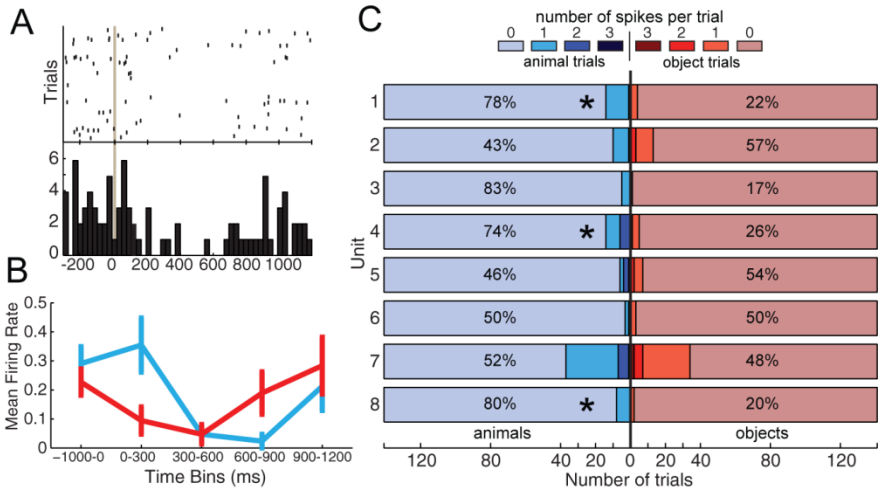

FIGURE III-6: PERIRHINAL CORTEX SINGLE UNIT FIRING RATES SHOW ANIMAL/OBJECT INFORMATION SPECIFICITY

A) Single unit raster-plot and peri-stimulus time histogram for a representative unit. B) Mean firing rate in five time bins for the same unit shown in (A) for animals (blue) and objects (red). From 0 to 300ms, the drop in firing rate for objects is much larger than the drop in response to animals. C) Number of spikes per trial (sorted into animal and object trials) for each of 8 identified units. Percentages indicate the proportion of trials with at least one spike in which the stimulus was a word associated with an animal (blue) or manmade object (red). Stars indicate the three units with statistically significant differences in firing between animal and objects trials (Wilcoxon rank-sum, $p<0.01$).

4. DISCUSSION

While many studies have demonstrated category-specific hemodynamic activity to images in posterior ventral temporal areas, responses to words in these areas have been more variable, and little has been seen more anteriorly. We report focal electrophysiological responses selective for words referring to animals versus objects in the inferotemporal, perirhinal and entorhinal sectors of the human anteroventral temporal lobe (avTL). Differences were observed both in measures sensitive to synaptic activity (LFP, gamma-band power, and CSD), and to unit-firing (MUA and single units) at multiple spatial scales. The timing, laminar location, and task correlates of this activity have implications for the mechanisms whereby more posterior ventral visual regions may show similar differential activation to the same stimuli. The avTL categorical responses may also contribute to stimulus-selective cuing of the hippocampal formation for recall, and of the amygdala for emotional evaluation. More generally, these findings provide additional evidence for a key role of avTL in semantic encoding.

Semantic category selectivity is present in the initial responses recorded in IT and PR at latencies as early as 130ms. Using CSD analysis (Ulbert et al., 2001, Einevoll et al., 2007, Pettersen et al., 2006), we identified the initial response as a sink in what was estimated to be middle cortical layers, the location where feed-forward afferents terminate (Van Hoesen and Pandya, 1975, Saleem et al., 1993, Saleem and Tanaka, 1996). These afferents are excitatory, as confirmed by the concurrent increase of category-selective multiunit activity. The principle source of feed-forward afferents to these structures in macaques arise largely in ventral occipitotemporal cortex (Mishkin et al., 1983, Suzuki, 1996, Suzuki and Amaral, 1994, Lavenex and Amaral, 2000, Martin-Elkins and Horel, 1992, Desimone et al., 1980). In humans these structures could correspond to the various high-level visual material-specific processors that

generally show their first peak of activity between 150-200ms and lie just anterior to classical retinotopic cortical areas (Allison et al., 1994, Allison et al., 1999, VanRullen and Thorpe, 2001, Halgren et al., 1999).

Indeed, it is possible that these afferents arise in the ventral occipitotemporal regions that respond selectivity to pictures of objects and animals (Liu et al., 2009, Chao et al., 1999, Chao and Martin, 2000, Devlin et al., 2005, Perani et al., 1999, Noppeney et al., 2006). However, we consider this possibility unlikely because these occipitotemporal areas do not reliably respond to words referring to these categories, but rather, their response is task dependent (Devlin et al., 2005, Mummery et al., 1998, Phillips et al., 2002, Price et al., 2003a).

In contrast, the category-selective responses to words reported here were present regardless of the task, including size, familiarity, and abstract/concrete judgment tasks. In fact, the word-memory task, which does not require explicit activation of an object's visual form, also yielded category-specific responses in these areas. Rather, our data suggest that avTL projections to ventral occipitotemporal cortex may cause it to display category-selective hemodynamic responses to words. Strong feedback projections between homologous areas have been demonstrated in macaques (Lavenex et al., 2002, Van Hoesen, 1982, Suzuki et al., 2000, Halgren et al., 1999). This hypothesis posits that occipitotemporal areas encode visual structural, rather than supramodal semantic information, resulting in automatic bottom-up activation by images, consistent with the early latencies reported by Liu et al. (2009). However, category-specific activation to words would only be observed in this area during tasks which required a full instantiation of that item's structural form. This interpretation is consistent with that proposed previously by Devlin et al. (2005) based on fMRI and neuropsychological results. MEG studies have also shown that more anterior areas in the

ventral stream provide feedback to ventral occipitotemporal areas after first pass processing of pictures to participate in the successful identification of visual objects (Bar et al., 2006), especially when precision is required (Clarke et al., 2011). Feedback projections arise in infragranular pyramidal cells in deep cortical layers. The current work recorded sustained activity in deep layers of avTL sites, also selective for semantic category, immediately after the feedforward peak in putative layer IV. Thus, the results of this chapter demonstrate category-specific synaptic and unit-activity in input layers at early latencies reflecting feedforward activation, and in deep layers at longer latencies reflecting the presumed source of feedback to ventral occipitotemporal areas. The large animal-object SVM weights seen in Chapter II at posterior temporal-occipital sensors around 400ms may be an extracranial manifestation of this feedback processing in visual structural areas.

Figure III-7 illustrates this model of category-selective perceptual and semantic information flow in the temporal lobe. The implication that activation of perceptual processing areas by words is secondary to lexico-semantic encoding, as well as being non-obligatory and task-dependent, may be inconsistent with some of the stronger claims of embodied cognition (Martin, 2007, Mahon and Caramazza, 2009). In our model, semantic category responses in avTL would reflect projections from the 'visual word form area' (VWFA) in the fusiform gyrus at the occipitotemporal junction (Dehaene et al., 2005, Crone et al., 2001b, Cohen et al., 2000, Halgren et al., 1994), and a possibly homologous auditory area in the superior temporal sulcus (Parker et al., 2005, Saur et al., 2008, Scott et al., 2000). It may be possible to conceive of the category-selective responses reported here as a continuation of progressively greater abstraction, a general theme of the ventral stream (Mishkin et al., 1983, Ungerleider and Haxby, 1994, Vinckier et al., 2007, Mesulam, 1998).

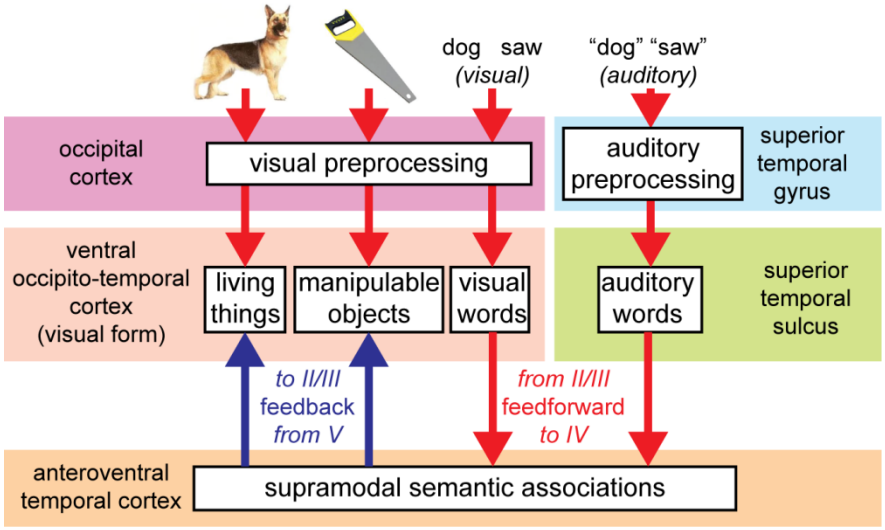

FIGURE III-7: MODEL OF LEXICO-SEMANTIC INFORMATION FLOW IN THE TEMPORAL LOBE

Visual inputs (either pictures or written words) are preprocessed by low-level occipital areas. Visual information proceeds to material-selective visual form areas in ventral occipito-temporal cortex that represent the structural information present in an image, or the orthographic representation of a written word. Category-specificity is possibly seen in this area to images due to the structural differences between living and non-living objects. This information then proceeds to anteroventral temporal cortex in which lexico-semantic associations are processed. Spoken word information proceeds along a similar pathway within the superior temporal cortices. When the particular task requires access of visuo-structural information after a written or auditory word input is perceived, feedback pathways (blue arrows) activate ventral occipito-temporal cortices.

In several sites, semantic category-selective responses were evoked by both visual and auditory words. This implies that there may also be input to avTL from auditory areas analogous to the VWFA, consistent with projections from the superior temporal lobe to this region in macaques (Saleem et al., 2000, Seltzer and Pandya, 1978), MEG co-localization of N400 responses to auditory and visual words in avTL (Marinkovic et al., 2003), and activation of anterior temporal lobe (aTL) to written and spoken language in fMRI (Spitsyna et al., 2006). Unfortunately, we did not record responses to auditory words from the laminar microarrays, and so could not determine if these responses were feedback or associative. In addition, the finding that images also evoke category-specific gamma responses in these same electrodes at approximately the same latency, or slightly later, is strong evidence that the information seen

in avTL is not visual-structural information. In this case, our model would predict that this picture information activates these visual-spatial representations in posterior occipito-temporal areas before amodal semantic information is sent anteriorly in a feedforward, rather than feedback, manner.

These supramodal category-selective responses to words are consistent with proposals that the avTL plays a central role in semantic representations (Lambon Ralph et al., 2010). Anomias and semantic dementia can be caused by lesions of this area (Bozeat et al., 2000, Davies et al., 2004, Damasio et al., 2004, Patterson et al., 2007, Jefferies et al., 2009, Mion et al., 2010), and the main generator of the N400, an event-related potential associated with lexico-semantic associations, is found here (Smith et al., 1986). Neuroimaging studies have often failed to find responses in aTL due to susceptibility artifacts or limited field-of-view (Visser et al., 2010b), however recent studies using distortion-corrected fMRI (Binney et al., 2010, Visser et al., 2010a), and rTMS (Lambon Ralph et al., 2009, Pobric et al., 2010a, Pobric et al., 2007, Pobric et al., 2010b) have provided further evidence for the importance of aTL in semantic processing. Interestingly, these studies have shown category-general semantic processing in lateral aTL (Pobric et al., 2010a, Lambon Ralph et al., 2010), while the results presented here demonstrate category-specific effects in more inferior and medial areas. This is consistent with the finding that semantic dementia patients generally do not show a category-specific deficit, however herpes simplex virus encephalitis patients, who have significantly greater medial involvement, do show such an effect (Lambon Ralph et al., 2007, Noppeney et al., 2007).

Our recordings demonstrated strong modulation of gamma-band activity by category membership when retrieving knowledge about the objects or animals represented by words. Gamma-band power from 30-40Hz, recorded extracranially, has previously been associated

with feature binding and the semantic lookup of lexical items (Pulvermuller et al., 1996a, Pulvermuller et al., 1996b, Lutzenberger et al., 1994, Tallon-Baudry and Bertrand, 1999). The data presented here are broadly consistent with a role for gamma activity in the semantic encoding of lexical items within the avTL. Our results also suggest that gamma-band activity tends to be more focal than low-frequency LFP activity, as others have proposed (Linden et al., 2010).

The anteroventral, inferotemporal and perirhinal areas showing early semantic category-selective responses project strongly to entorhinal cortex, the gateway to the hippocampus (Insausti et al., 1987, Burwell, 2000). O'Keefe and Nadel (1987) originally proposed that the human hippocampus maps semantic space in a manner analogous to the rodent mapping of physical space. Indeed, human hippocampal neurons selectively fire to specific words (Heit et al., 1988) which may correspond to the firing of rodent hippocampal neurons to specific places. More recently, the apparent raw material for constructing place cells has been identified as the grid cells of EC in rats (Hafting et al., 2005). In a similar way, the firing of human entorhinal cells to specific semantic categories may provide the inputs used by hippocampal cells to select for individual words (Kreiman et al., 2000, Heit et al., 1988).

In this chapter, we have also seen that SVMs allow for the robust decoding of semantic category information from microelectrode recordings. In fact, because of the increased signal-to-noise ratio of the microelectrode recordings, we are able to achieve much higher accuracies when decoding from single trials (~75% for microelectrodes versus ~60% for MEG). Similar to the improved accuracy when combining EEG and MEG features, an improvement is also observed when combining CSD and MUA features. In this case, while CSD and MUA and highly correlated, they carry slightly different information. MUA contains information regarding that

output of neurons that, by definition, is super-threshold excitation. CSD, on the other hand, reflect the depolarization of this population of cells and may contain a great deal of sub-threshold information that is not necessarily transmitted further. Therefore, the greater information content of the MUA signal is consistent with the idea that the firing rate of neurons is the main method of communication between cells and brain areas. However, the CSD may contribute additional information because it reflects the firing rates of neurons that are upstream from the population in question, or simply because it provides another correlated measurement that, in a sense, helps increase the signal-to-noise ratio.

Interestingly, the decoding using sliding windows demonstrated two peaks of increased information in the CSD/MUA features 150-350ms and 500-600ms. While this early increase in accuracy likely reflects the first-pass semantic information coming into perirhinal, entorhinal, or inferotemporal cortices, the origin of the later peak is less clear. It is possible that this later activity, that roughly corresponds to the CSD activity in the upper layers, is due to recurrent processing within these areas or feedback information from other areas.

The results of this chapter demonstrate not only that category selectivity is present in avTL, but that this selectivity is present on the first pass of activity through this area. This activity is seen in measures sensitive to both synaptic and unit-firing activity at multiple spatial scales. The model proposed here suggests that avTL encodes semantic categories and provides this information to posterior ventral temporal areas when task demands so require, resulting in their variable category-selective hemodynamic response to words. In addition, SVMs are able to robustly extract this information from the microelectrode recordings and verify that semantic information occurs early. The results presented in this chapter are consistent with the idea of an anterior temporal semantic hub that coordinates the modality specific

information that is distributed across the cortex, although further research is necessary to support this claim.

While the results in this chapter suggest that both auditory and visual information leads to amodal semantic information in avTL, whether the lexical inputs to this system are also amodal, or instead modality-specific, is unclear. One may imagine that visual and auditory words have separate lexical representations that directly feed into this semantic system. At the other extreme, auditory and visual information may need to activate a completely separate supramodal word representation before semantic information may be retrieved. A number of intermediate possibilities also exist, but further study is necessary for distinguishing between these options. Regardless, it is important to understand how low-level auditory and visual word stimuli are processed for lexical recognition and eventual input into the semantic system. Similar to the ventral visual object recognition pathway that maps visual words or images to meaning, potentially passing through the VWFA, the speech perception system is hypothesized to have an anteroventral "what" pathway that maps sounds to meaning (Hickok and Poeppel, 2007). In the next chapter, we will study the role of one stage of this pathway in the processing of speech stimuli.

IV. SPEECH-SPECIFIC TUNING OF SINGLE NEURONS IN HUMAN TEMPORAL LOBE

1. INTRODUCTION

While Chapters II and III examined the neural basis of semantic information, it is equally important to understand how low-level visual and auditory information, in the form of written and spoken words, leads to lexical representations and eventually interfaces with the semantic knowledge system. The high frequency of obtaining temporal lobe coverage in epilepsy patients with implanted electrodes makes the study of the auditory word processing pathway potentially more fruitful than examination of visual word processing in ventral occipital regions. Understanding the fundamental acoustic/phonetic building blocks the brain uses to construct words is of crucial importance to understanding speech processing, however little regarding these features is known. Many important questions regarding the processing of speech can be raised. For example, despite the recent evolutionary origin of language, are there neurons in the human brain that are tuned exclusively to speech that fail to respond to non-speech sounds that are equally complex in structure (e.g. environmental sounds)? If so, are the receptive fields of such neurons best described in terms of phonemes or other linguistic constructs? Is there evidence of columnar organization in speech-sensitive neurons? Do such neurons respond to written words in a manner that reflects activation of phonological information? Creutzfeldt et al. (1989a) have demonstrated that neurons in lateral temporal lobe respond to speech, however, many of these important questions are left unanswered.

The neural processing steps used to recognize words from low-level acoustic features remain to be specified. It is widely accepted that auditory processing is hierarchical in nature where relatively simple features are progressively combined to build more complex representations in downstream areas (Hickok and Poeppel, 2007, Rauschecker and Scott,

2009). This pathway begins with tonotopic frequency tuning in primary auditory cortex (Bitterman et al., 2008, Howard et al., 1996), and in primates and other animals, it has been shown that downstream neurons are tuned to frequency-modulated sweeps and other complex spectrotemporal representations (Rauschecker, 1998, Rauschecker and Scott, 2009). In humans, neuroimaging studies have suggested that this hierarchy applies to speech stimuli (Binder et al., 2000, Wessinger et al., 2001), however, hemodynamic activation is a population measure that does not determine the receptive fields of single neurons. Further, the temporal resolution of hemodynamic activations does not permit conclusions regarding their temporal sequence, and thus the observed spatial hierarchy may not necessarily reflect sequential processing, but rather feedback or interactive activity.

In addition, multiple hypotheses exist for the neuroanatomical organization of the speech processing stream. The traditional view suggests that the posterior superior temporal cortex, near Wernicke's area, is the main region involved in the representation of speech sounds (Steinschneider et al., 2011, Desai et al., 2008, Chang et al., 2010, Boatman et al., 1995, Crone et al., 2001a, Geschwind and Levitsky, 1968, Wernicke, 1874). There is growing evidence, however, that anterior superior temporal cortex is important for phonetic processing of speech and is part of the auditory "what" stream, involved in mapping auditory inputs to meaning (Obleser et al., 2006, Obleser et al., 2010, Arnott et al., 2004, Zatorre et al., 2004, Binder et al., 2004, Scott et al., 2000, Warren et al., 2006, Scott et al., 2006).

This chapter is a detailed examination of the auditory and language responses of a large number of single-units recorded from the left anterior superior temporal gyrus (aSTG) of a 31 year old, right-handed male patient with epilepsy. A 96 microelectrode array allowed for the recoding of single-unit extracellular action potentials in layer III/IV of auditory-responsive

cortex. The patient performed a large battery of auditory and language tasks that required semantic (size) judgments of auditory words referring to objects or animals (SA), comparison of spoken words to vocoded-speech during picture-word matching (WN), passive listening to pure tones, environmental sounds, and time-reversed words, repetition of a set of spoken words, and participation in spontaneous conversation. The patient also performed the visual counterpart of the SA task using written (instead of auditory) words (SV).

2. METHODS

2.1. PARTICIPANT

The patient was a 31 year old right-handed male with medically intractable epilepsy who was admitted to Massachusetts General Hospital for semi-chronic electrode implantation for surgical evaluation. The patient was left-hemisphere language dominant based on a WADA test, and was a native English speaker with normal hearing, vision, and intelligence. His seizures were partial complex in nature and typically began in mesial temporal depth electrode contacts. Resected tissue included left anterior temporal lobe (including the site of the microelectrode implantation), left parahippocampal gyrus, left hippocampus, and left amygdala. The patient was seizure-free at ten months post-resection. The patient gave informed consent and was enrolled in this research under the auspices of Massachusetts General Hospital IRB oversight in accordance with the declaration of Helsinki.

2.2. ELECTRODES AND RECORDING

A microelectrode array (Blackrock Microsystems, Salt Lake City, UT), capable of recording the action potentials of single units, was implanted in left anterior superior temporal gyrus. This 4x4mm array consists of 100 (96 active) penetrating electrodes, each 1.5mm in

length with a 20μm exposed platinum tip, spaced 400μm apart. Recordings were obtained by a Blackrock NeuroPort data acquisition system at 30kHz with bandpass filtering from 0.3Hz to 7.5kHz. The decision to implant the array in the superior temporal gyrus was based on clinical considerations; this was a region that was within the expected resection area and was, in fact, resected upon completion of the intracranial EEG (iEEG) investigation. The region surrounding the array was removed *en bloc* and submitted for histological processing. Staining with hematoxylin and eosin revealed that the tips of the electrodes were at the bottom of cortical layer III, close to layer IV, and that the surrounding cortical tissue was normal.

In addition to this microelectrode, clinical intracranial macroelectrodes were implanted based on clinical reasoning alone, and covered a large portion of left lateral cortex including frontal, temporal, and anterior parietal areas. Electrodes consisted of one 8x8 grid of subdural macroelectrode contacts (spaced 1cm apart), two 4-contact macroelectrode strips covering the left posterior-inferior temporal lobe, one 4-contact strip covering the left orbito-frontal region, and two 4-contact strips over the left frontal pole (Adtech Medical, Racine, WI). iEEG was continuously recorded from these clinical electrodes at 500Hz with bandpass filtering from 0.1 to 200Hz. Electrodes were localized using a volumetric coregistration algorithm to align the anatomical information of the pre-operative MRI with the electrode localization in the post-operative CT (Dykstra et al., 2012).

2.3. AUDITORY TASKS

The patient performed a number of auditory tasks designed to examine various aspects of speech and non-speech sound processing. The auditory size judgement task (SA), described in Section II.2.2, was performed in which the participant was presented with spoken words corresponding to various objects, and was asked to press a button for objects that were larger

than one foot in any dimension. Words were spoken by a male speaker, normalized in power and length (500ms), and presented with a 2200ms stimulus onset asynchrony (SOA). 800 randomly ordered trials were evenly split between novel words presented only once for the entire experiment (400 trials), and repeated words which consisted of a set of 10 words repeated 40 times each. The ten repeated words were "claw," "cricket," "flag," "fork," "lion," "medal," "oyster," "serpent," "shelf," and "shirt." Half of the trials required a button press yielding a 2x2 balanced design. Sounds were presented binaurally using Etymotic ER-1 earphones (Elk Grove Village, IL). A visual version of the SA task was also performed (SV) as described in Section II.2.2.

A word-noise task (WN) was performed in which an object picture was presented followed by a spoken word or noise. The picture (<5% visual angle) appeared for the entire trial duration of 1300ms, and the auditory stimulus, either a congruously or incongruously paired word or noise stimulus, was presented binaurally 500ms after picture onset. Four conditions were presented in random order: picture matched-words, picture matched-noise, picture mismatched-words, picture mismatched-noise. The participant was asked to press a button to matches. Words were single-syllable nouns recorded by a female native speaker. Noise stimuli were band-passed and amplitude-modulated white noise made to match the acoustic structure of a corresponding word. The power in each of 20 equal bands from 50-5000 Hz, and the exact time versus power waveform for 50-247, 248-495Hz and 496-5000Hz were matched between the noise and word stimuli (Shannon et al., 1995). Sounds (mean duration= 445±63ms; range= 304-637ms; 44.1kHz; normalized to 65dB average intensity) were presented binaurally through Etymotic ER-1 earphones.

Several sets of other sounds were also presented to the patient. 7.2 second sequences of randomly selected pure tones were presented to explore responses to simpler acoustic stimuli. Tones were 100ms in length including 10ms raised cosine on and off ramps and were centered at 0.239, 0.286, 0.343, 0.409, 0.489, 0.585, 0.699, 0.836, 1, 1.196, 1.430, 1.710, 2.045, 2.445, 2.924, 3.497, 4.181, or 5.000kHz. Tones were placed randomly in time and frequency within each band with an average within-band stimulus-onset asynchrony (SOA) of 800ms (range: 100-1,500 ms). Within each band, the exact frequency of any given tone was within an estimated equivalent rectangular bandwidth (ERB), where ERB=24.7*(4.37*fc+1), where fc is the center frequency of a given band, in kHz.

In addition to the pure tones, the participant was presented the SA word stimuli and asked to repeat them out loud. The subject began speaking, on average, 411±119ms after the end of the stimulus, and the SOA was 3000ms. The auditory word stimuli from the SA task were also time-reversed and presented to the participant who was asked to passively listen. Words were presented with the same 2200ms SOA as the SA task. Time reversal of words preserves the spectral content of the stimuli but changes the temporal structure of the sounds. A set of environmental sounds, natural (e.g. birds chirping, waterfall) and manmade (e.g. clapping hands, breaking glass), were also presenting to the patient. A total of 30 environmental sounds were presented 5 times each in pseudorandom order with a 3500ms SOA. Finally, a spontaneous conversation between the patient and researchers was recorded using a far-field microphone. The entire conversation was manually transcribed and all word-boundaries were marked.

2.4. SPIKE SORTING AND ANALYSIS

To extract spikes from the microelectrode recordings, continuous data was highpass filtered at 250Hz using a 6^{th}-order Bessel filter, and an amplitude threshold of 4 standard deviations was used to choose action potential waveforms. Extracted spikes were manually sorted using Offline Sorter (Plexon, Dallas, TX) in various feature spaces including principal components, peak-valley amplitude, and non-linear energy. To allow for characterization of units that were present across experiments, multiple sessions were concatenated and sorted together. Units were characterized as single or multi-units based on quality of sorted clusters and amplitude of waveforms. For all subsequent analyses, both single-units and multi-units were included. Putative inhibitory interneurons were identified based on waveform shape, full-width at half max, and valley-to-peak time (Bartho et al., 2004, Peyrache et al., 2012).

To statistically determine and quantify the magnitude and latency of unit responses to particular stimuli, a non-parametric cluster-based statistical test was used. Similar to the non-parametric test utilized for testing continuous LFP and HGP data in Section III.3.1 (Maris and Oostenveld, 2007), statistics were computed for individual bins of peri-stimulus time histograms (PSTHs). Specifically, the binned firing rates of a baseline period from -300 to 0ms before the stimulus are compared to each bin after stimulus onset using a two-sided T-test. Clusters of consecutive bins with bin-level significance of $p_{bin}<0.05$ are found, and the summed T-statistic of each cluster is computed. The null distribution of cluster-level statistics is computed by randomly permuting the bins in time and recomputing clusters 1000 times. The cluster-level summed T statistics are compared to the null distribution and a cluster is deemed significant if the cluster level probability was $p_{cluster}<0.05$. The earliest bin in a statistically significant cluster is taken to be the response latency of that particular unit, and the average firing rate within that cluster is taken to be the magnitude of the response.

To quantify responses to phonemes, all phoneme boundaries were manually marked for all relevant stimuli. Formants were computed using Wavesurfer (http://www.speech.kth.se/wavesurfer/) using 20ms windows overlapping by 10ms. For each vowel phoneme, the midpoint was chosen to determine the characteristic formants for each vowel. For analysis of phonological information in written words, the Carnegie Mellon University Pronouncing Dictionary was used to obtain phonetic transcriptions of words in the SV task (http://www.speech.cs.cmu.edu/cgi-bin/cmudict). To obtain word frequency information, words from the SV task were looked up in the HAL corpus (Lund and Burgess, 1996), resulting in a mean word frequency of 2849, median of 1422, and range of 2 to 37,798. To group words into two frequency classes, words below the median value of 1422 were grouped into the low-frequency class, with words above this median taken as the high-frequency class.

2.5. SPECTROTEMPORAL RECEPTIVE FIELD ESTIMATION

To compute the spectrotemporal receptive fields of each unit, two sets of spectral features of the stimuli were used: power in linearly-spaced frequencies from 50Hz to 4kHz, and Mel-frequency cepstral coefficients. Features were computed in 20ms windows overlapped by 10ms for all stimuli in the SA and WN tasks. To compute the Mel-frequency cepstral coefficients, melfcc MATLAB function was used to compute the cepstral coefficients between 200Hz and 8kHz using a liftering exponent of 22. The first 13 coefficients were extracted, and an additional broadband energy term, computed in 20ms windows shifted by 10ms, was concatenated to form the final feature vector. MFCCs and band-power were computed for all auditory stimuli in the SA and time-reversed word tasks.

The method described by Theunissen et al. (2001) was used to estimate the STRFs for each unit. This method compensates for correlations within the stimuli to generate the optimal linear filter, or impulse response, which best characterizes the relationship between stimulus features and firing rate. To predict the firing rate of these units to words, firing rate was smoothed using a 120ms Gaussian kernel, and the STRFs were computed using the novel words in the SA task. The resulting impulse response was convolved with the time-course of stimulus features generated for the repeated words, yielding a predicted PSTH.

More specifically, if the impulse response, or STRF, of a given neuron is time-invariant, we can predict the response, $\hat{r}$, of the neuron to a given stimulus by,

$$\hat{r}[t] = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} h[i,j]s[t-i,j]$$

where h is the impulse response, s is the stimulus, index i is summed over N points in time, and j is summed over M spectral features. If we combine the temporal and spectral indices, we can write this as $\hat{r}[t] = \boldsymbol{h}^{\mathrm{T}} s_t$. By minimizing $\langle(\hat{r} - r)^2\rangle$, the solution for h becomes $\boldsymbol{h} = \boldsymbol{C}_{ss}^{-1}\boldsymbol{C}_{sr}$, where $\boldsymbol{C}_{ss}$ is the stimulus autocorrelation matrix, $\boldsymbol{C}_{sr}$ is the cross-correlation vector between the stimulus, s, and the response, r.

2.6. DECODING

To decode either repeated words or phonemes, a set of features was computed from the unit firing rates for each trial. For the repeated word classification task, all units which demonstrated a statistically significant response to auditory words were used. For each unit, a time window was computed in which the unit's firing rate significantly changed from baseline. Subsequently, for each trial, the number of spikes occurring within each unit's chosen time

window was used as one of the features for the classifier. For phoneme decoding, this window was fixed from 50 to 250ms after phoneme onset.

To examine changes in information over time, either sliding windows or cumulative windows were used to compute firing rates. For sliding windows, the firing was computed in a 50ms window for each unit beginning at 0ms (the window therefore covered 0-50ms) and after decoding the information in this window, the window was shifted 10ms forward. For the cumulative window analysis, 25ms windows were used, and instead of shifting the window, subsequent non-overlapping windows were concatenated to the growing feature vector between each round of decoding. This allowed for analysis of information in a time frame of 0-25ms up to 0-1000ms.

A Naïve Bayes Classifier (as discussed in Section I.3) was used to decode word or phoneme-specific information from the computed features. The Naïve Bayes Classifier assumes that all features, $f_1, f_2, \ldots, f_n$, are independent which makes joint probabilities simply the product of the marginal probabilities. Because the number of features used in the classification of these data was smaller than the number used in Chapter II, the choice of the Naïve Bayes classifier over SVM is reasonable. Furthermore, because spike trains can naturally be treated as Poisson processes, using a Poisson distribution for modeling the probabilities was logical, intuitive, and simple. To train the classifier, for each combination of unit and class, a Poisson distribution of spike counts was estimated (via maximum likelihood) allowing for computation of probabilities of observed firing rates. Accuracies were estimated via 10-fold cross-validation.

3. RESULTS

3.1. SINGLE UNIT SORTING RESULTS

Units were manually sorted for each of the seven experiments. Because experiments were performed over the course of 3 days, identified units varied from task to task. For analyses where units were compared between experiments, sorting was performed simultaneously over all tasks of interest. From the SA task, a total of 142 units were identified with 58 units characterized as likely single-units and 84 potential multi-units. A total of 146 units were identified from the WN task (63 single-units, 83 multi-units), 166 units during presentation of the time-reversed words (77 single-units, 89 multi-units), 144 units during presentation of pure tone sequences (77 single-units, 67 multi-units), 169 units during repetition of auditory words (79 single-units, 90 multi-units), 171 units during a spontaneous conversation (77 single-units, 94 multi-units), and 181 units during the SV task (86 single-units, 95 multi-units). Single-units demonstrated overall firing rates between 0.0004 and 11.1 spikes/sec (mean = 0.37 spikes/sec). 17% of the identified single-units were putative inhibitory interneurons based on waveform shape, full-width at half max, and valley-to-peak time (Bartho et al., 2004, Peyrache et al., 2012). The mean firing rate of these inhibitory cells was 1.96 spikes/sec as compared to the mean firing rate of excitatory single units at 0.17 spikes/sec. Responses were similar for single and multi-units, and for putative pyramidal cells and interneurons, and they are combined for subsequent analyses unless specifically indicated.

3.2. SINGLE UNIT WORD SPECIFICITY

The patient was first asked to listen through headphones to a set of recorded spoken words that corresponded to concrete objects or animals, and indicate if the item was larger than a foot in any dimension (SA task). Half of the trials involved 400 words that were only

presented once during the experiment while the other half involved 10 words that were presented 40 times each. It was immediately clear that many of the units present on the microelectrode array responded strongly to these spoken word stimuli. Offline analysis of these responses demonstrated that a large number of units not only responded to the auditory word stimuli in general, but responded very differently to specific words. Two examples of such units are shown in Figure IV-1. These two cells demonstrated bidirectional monosynaptic connectivity (based on cross correlation and waveform shapes, Figure IV-1B) and strong word-specific responses to the repeated words in the SA task. Interestingly, the excitatory cell demonstrated narrower tuning than the inhibitory cell (Figure IV-1c). Unit 6a, the putative inhibitory interneuron, demonstrated differences in firing rate to the 10 repeated words with the largest responses being to the words "claw," "cricket," and "oyster." For unit 6b, the putative pyramidal cell, while "claw" and "cricket" also evoked large responses, "oyster" did not.

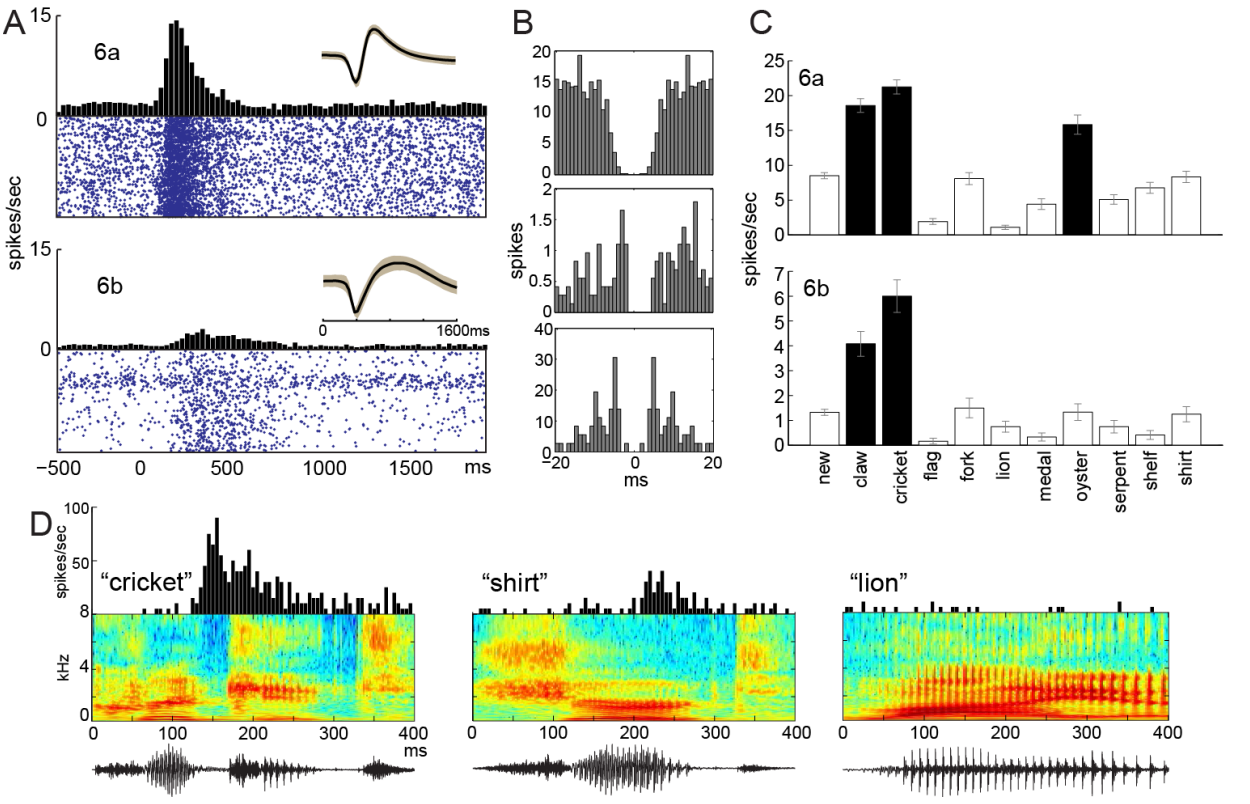

FIGURE IV-1: UNITS DEMONSTRATE DIFFERENTIAL FIRING TO INDIVIDUAL WORDS.

A) PSTHs and raster plots of units 6a (top) and 6b (bottom) in response to word stimuli. B) Autocorrelograms for units 6a and 6b (top and bottom respectively), and crosscorrelogram for unit 6b in relation to unit 6a (middle) suggesting bidirectional, monosynaptic connectivity between an inhibitory interneuron (6a) and pyramidal cell (6b). C) Firing rates for units 6a and 6b to each of the 10 repeated words demonstrate robust word-specific firing. D) PSTHs for unit 6a in response to three example words with corresponding stimulus spectrogram and waveform plots below show differences in magnitude and latency of response.

In total, 66 of the 141 units exhibited a statistically significant response to auditory words in the SA task ($p<0.05$, Figure IV-2b). 59 units demonstrated an increase in firing to words while 7 showed a decrease in firing. Baseline firing rates of responsive units varied from 0 to 5.16 spikes/sec (mean=0.31 spikes/sec) with changes in firing rate from 0.03 to 12.4 spikes/sec (mean=0.75 spikes/sec). Peak firing rates ranged from 0.31 to 14.3 spikes/sec (mean = 0.50 spikes/sec). Response latencies varied from 20 to 940ms (mean = 308ms) after word onset. Of the identified units, 31 demonstrated differential firing to the 10 repeated

words in the SA task ($p<0.05$, Kruskal-Wallis, between 100-900ms). No cells responded differentially to words referring to animals versus manmade objects (like the units reported in Section III.3.5), or to novel versus repeated words.

While these units demonstrate differential firing to words, tuning to a number of different features may lead to this apparent word specificity. At the lowest level, it is possible that these units simply respond to a particular frequency or sound intensity that is present in a subset of the presented words. It is also possible that these units are responding to specific acoustic features of spoken words such as complex time-frequency components, combinations of formant frequencies, or even phoneme identity. At the highest levels, these units may encode the auditory representations of full words. We therefore tested the response of these units to a diverse set of auditory stimuli that spanned a wide range of acoustic complexity.

3.3. SPATIAL ORGANIZATION OF RESPONSES

The regular spacing of the 10x10 microelectrode array allowed the spatial organization of unit response properties to be examined. Although the identified units were fairly evenly distributed across the array (Figure IV-2A), those responding to the auditory word stimuli were clustered on one side of the array (Figure IV-2B). More specific properties such as response latency (Figure IV-2C) and word-specific response profiles also demonstrated spatial correlations across the array. The correlation of responsiveness to words extended to 800μm, the correlation for response latencies was significant to 600μm, and the maximum correlation distance for 10-word response profiles was 400μm (Figure IV-2D). Thus, more general response properties exhibit spatial correlations over larger distances than do more specific response properties. Therefore, similar to the spatial arrangement of frequencies in primary auditory cortex, high order auditory information may also be spatially organized.

This analysis shows that the response profiles across words were similar for multiple units recorded by a given contact, and slightly correlated with the profiles of neurons recorded at adjacent contacts. We also tested if the HGP responses at contacts showing correlated profiles showed significantly different response profiles to different words. This was found in 2 of 11 such electrodes, thus indicating that the specificity of local cells for particular words can also be observed at the population level.

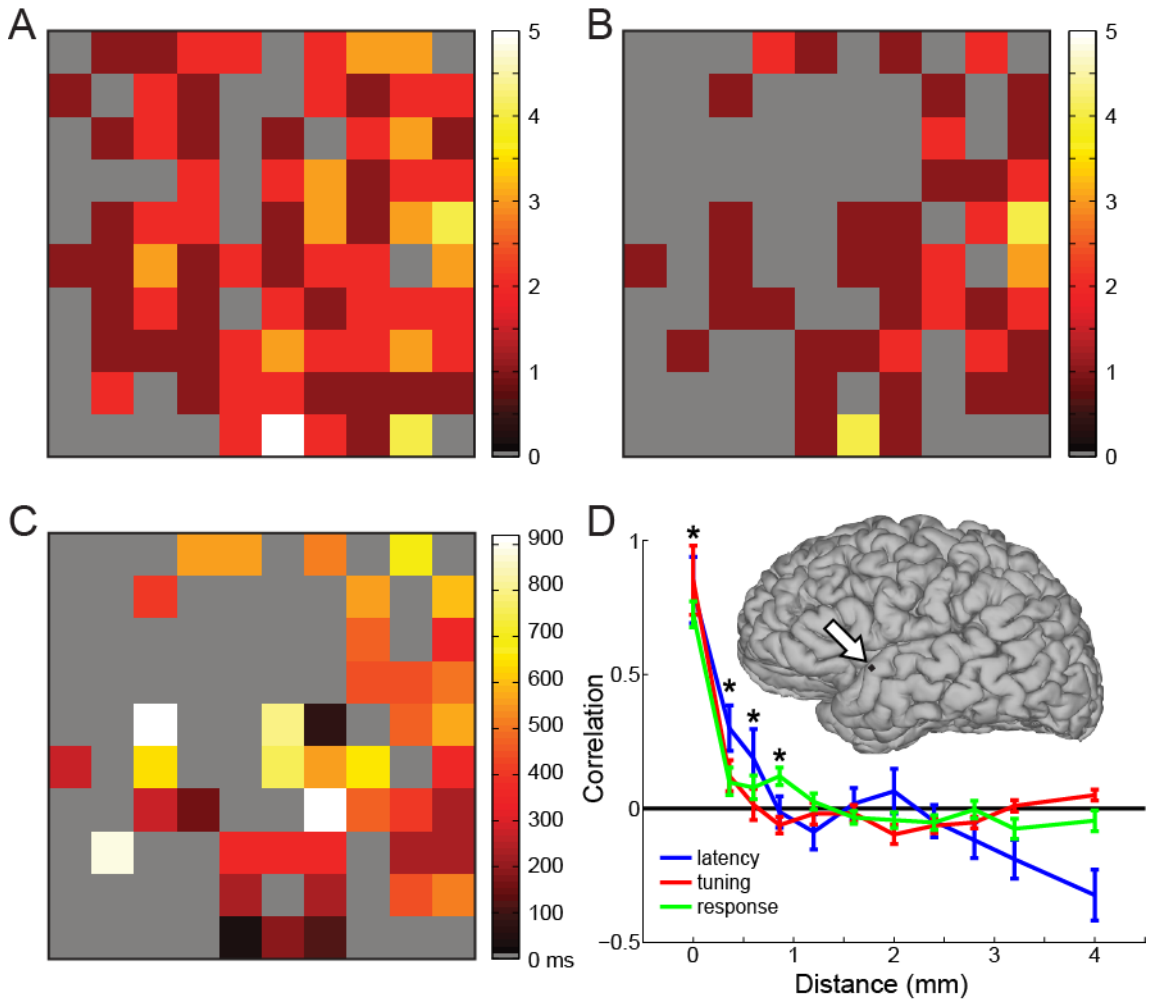

FIGURE IV-2: UNITS DISPLAY SPATIAL ORGANIZATION FOR RESPONSE AND TUNING PROPERTIES.

A) Number of identified units on each electrode of the array during the SA task, organized spatially. A total of 141 units were identified. B) The number of units on each electrode which demonstrated a statistically significant response to word stimuli greater than baseline. A total of 66 units responded to auditory word stimuli. C) The distribution of response latencies across each channel

(Figure IV-2 continued) of the array. The mean latency is shown for electrodes with more than one unit. D) Significant spatial correlation of any response to words was found up to 800μm (green), for response latency up to 600μm (blue) and for individual-word selective responses up to 400μm (red) (t-test, $p<0.05$). Inset shows the anterior STG location of the microelectrode array.

3.4. RESPONSES TO NON-SPEECH SOUNDS

In order to test whether these units responded to non-speech auditory stimuli, the patient was asked to passively listen to sequences of 100ms pure tones ranging from 240Hz to 5kHz. None of the units showed statistically significant responses to these simple sounds at any frequency ($p>0.05$, Figure IV-3A).

The patient also listened to auditory words and noise-vocoded speech that was matched with each word. The noise-vocoded stimuli contained the same time-course of power in 3 frequency bands as the matched word, but the fine-scale spectral information within these bands was replaced with amplitude modulated, band passed white noise. The task was to decide if the words matched a picture presented immediately prior. Three of the 60 units that responded to words also responded to noise-vocoded speech, however firing rates were significantly lower; on average signal-correlated noise elicited only 35% of the firing rate to words (Figure IV-3B). This suggests that neurons in this area largely respond to acoustic/phonetic features specific to words, and not to the amplitude envelope or the rough matching between band-power found in noise-vocoded speech.

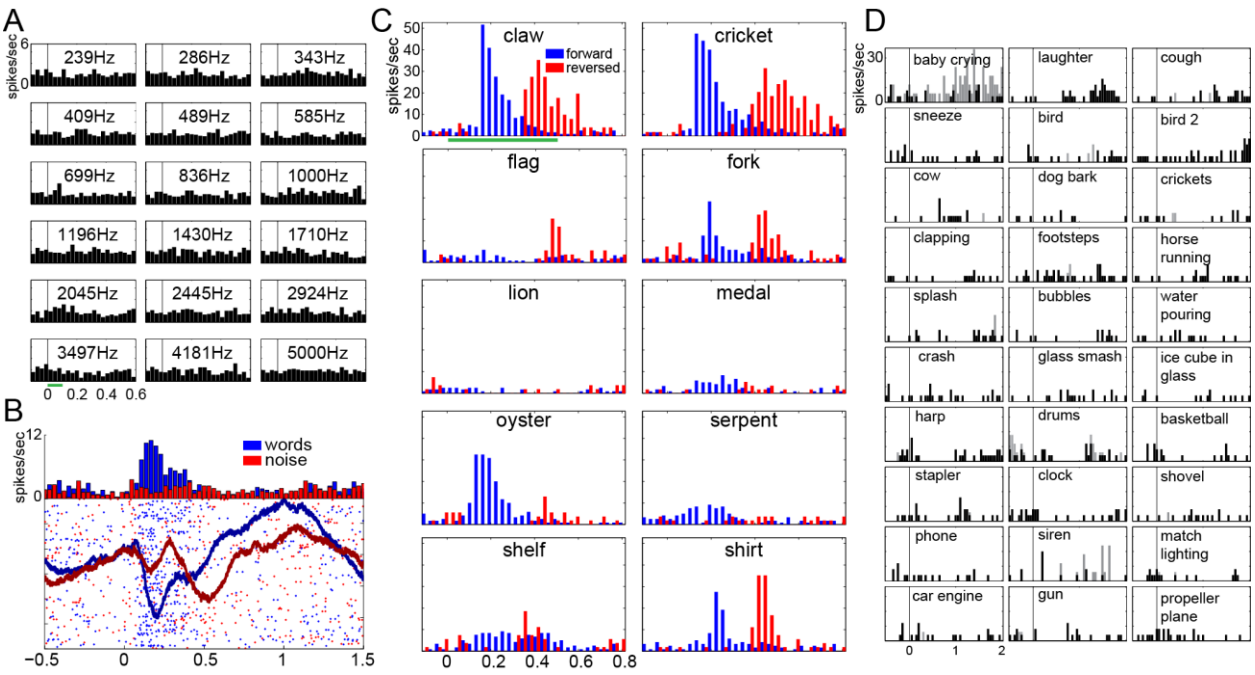

FIGURE IV-3: UNITS FAIL TO RESPOND TO NON-SPEECH SOUNDS.

A) PSTHs for unit 6a to 100ms pure tones at 18 different frequencies demonstrate no changes in firing rate. B) Unit 6a demonstrates a large response to auditory words but a much reduced response to word-matched noise. Local Field Potential waveforms are superimposed on the raster plot. C) PSTHs to each of the 10 repeated words for the SA task played forward (blue) or time-reversed (red) demonstrates shifted latencies and relatively reduced responses in most cases. D) Only unit 6a (black) and unit 36a (gray) demonstrated statistically significant responses to any of the presented environmental sounds (Monte-Carlo permutation test, $p<0.05$). Unit 6a only responded to laughter while unit 36a only responded to a baby crying, both human vocal sounds. Magnitude of responses was significantly lower than the response to spoken words.

The patient was then asked to passively listen to the same auditory words in the SA task that had been time-reversed. Time-reversed words contain the same frequency information and acoustic complexity as the original words, however the temporal structure of the frequency information is changed and many of the sounds are not-phonetically possible. Such stimuli have often been used as controls in hemodynamic studies of word responses (Howard et al., 1992, Perani et al., 1996, Price et al., 1996, Hirano et al., 1997, Binder et al., 2000, Crinion et al., 2003). In essence, vowel sounds are relatively preserved while consonants (especially stop consonants) are often distorted. A total of 17% of the identified units responded to time-

reversed words (as opposed to 47% for normal words), and the magnitude of the response was significantly smaller. The mean increase in firing for time reversed words was 0.21 spikes/sec (versus 0.75 spikes/sec for normal words). For several units, responses were also significantly delayed in latency (Figure IV-3C). The fact that some of these units respond to time-reversed words suggest that they are responding to complex auditory features rather than full words, however the smaller magnitude of the response suggests that they are tuned to meaningful speech sounds.

While these control stimuli elicit relatively smaller responses than spoken words, they are artificially constructed synthetic sounds. It is possible that the identified units respond equally well to naturally occurring environmental sounds. A set of 30 environmental sounds, both man-made and natural, were presented to the subject. Only two of the identified units demonstrated statistically significant responses to any of the stimuli (Figure IV-3D). Unit 6a only responded to male laughter while unit 36b only responded to a baby crying. Interestingly, both of these stimuli are human vocalizations.

3.5. RESPONSES TO PHONEMES

To explore what property of spoken words these units responded to, the first and second formants (F_1 and F_2) of all stimulus words were computed, and phoneme boundaries marked and labeled manually using Wavesurfer (http://www.speech.kth.se/wavesurfer/). Formant values at the midpoint of each vowel was correlated to the firing in response to that vowel from 50-250ms for all units. Only unit 6a demonstrated significant correlations, with a moderate negative F_1 correlation of $\rho=-0.13$ (Spearman, $p<0.01$) and positive F_2 correlation of $\rho=0.11$ (Spearman, $p<0.05$).

Peri-stimulus time histograms (PSTHs) were generated for each of the phonemes present in the SA and WN stimuli. While the firing rates of many units were too low to confidently infer a pattern of response across different phonemes, unit 6a clearly showed phoneme specific firing to several vowel sounds beginning at approximately 70ms and peaking at 100ms (Figure IV-4). These phonemes included the high-front vowels [ɪ], [i], [oɪ], [oʊ], and [u]. Several consonants, such as [p], [b], [t] and [f] also demonstrate increases in firing around 100ms. Because the probability of a particular phoneme occurring in a specific location in a word is not independent from the surrounding phonemes (i.e. phonemes have significant correlations within words), some of the phoneme PSTHs, such as [ŋ], demonstrate increases in firing before 0ms. This is likely due to the high probability of occurrence of specific preceding vowel-sounds (e.g. [ɪ] as in words ending with "ing"). Overall, 22 units demonstrated significant responses to at least one phoneme, with 16 of the 22 units responding to more than one phoneme (Figure IV-4C). Furthermore, while more units exhibited significant responses to consonants (Figure IV-4D), 9 of the 22 units responded to both consonants and vowels.

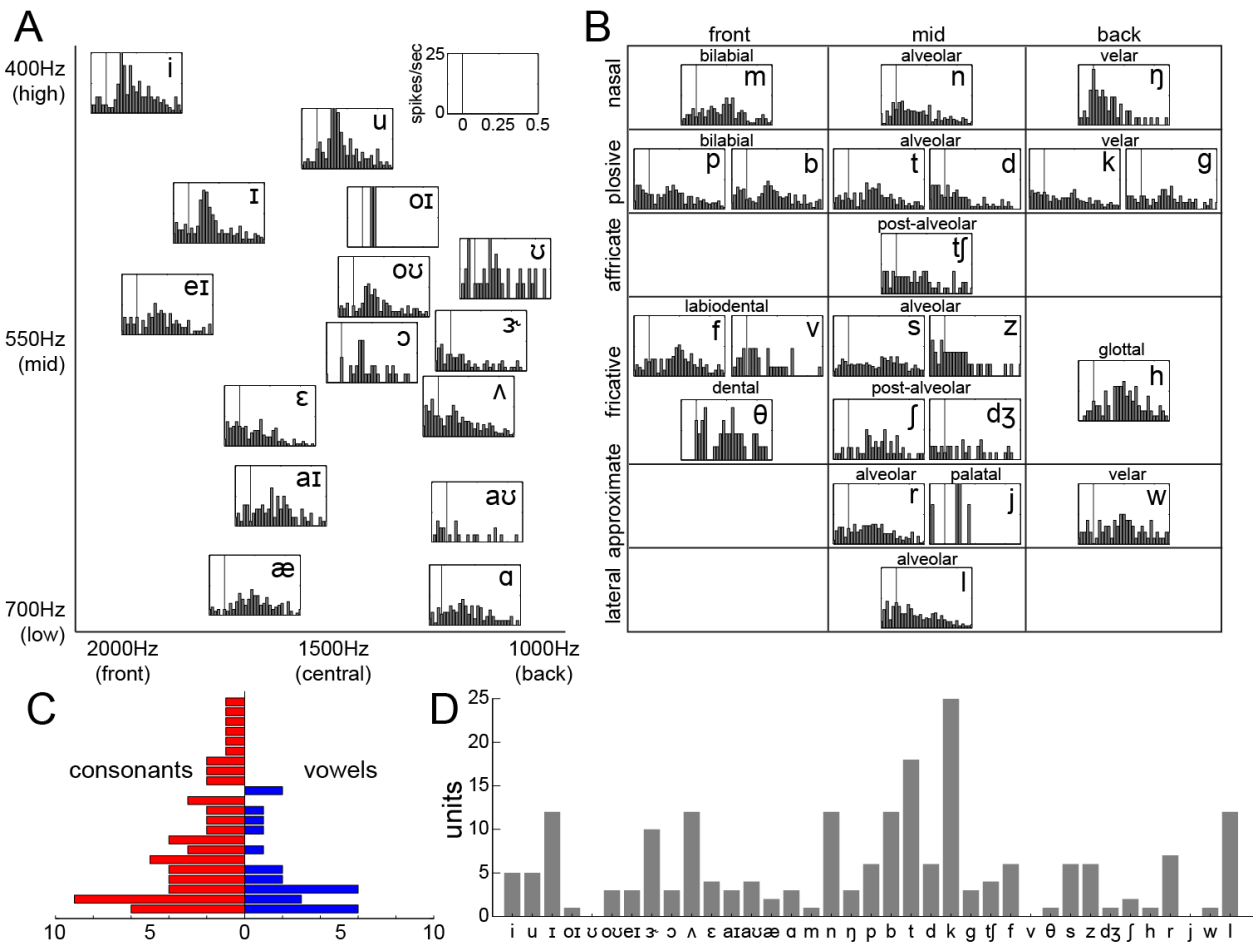

FIGURE IV-4: PSTHS DEMONSTRATE FIRING TO A SUBSET OF PHONEMES.

A) PSTHs for unit 6a for each vowel phoneme approximately arranged in formant space. B) PSTHs for each consonant. C) Number of consonants or vowels each unit significantly responded to relative to baseline ($p<0.05$). Each horizontal line represents a different unit. Bars with an adjacent black dot indicate single (versus multi) units. D) Distribution of units responding to each phoneme with vowels on the left and consonants on the right.

The fact that many of these units respond to a number of different phonemes suggests that they may be tuned to particular combinations of consonants and vowels. In fact, many previous studies utilize consonant-vowel (CV) pairs as stimuli for examining phonological representations, although the basis for doing so is unclear (Jancke et al., 2002, Liebenthal et al., 2005, Obleser et al., 2010, Steinschneider et al., 2011). By grouping vowels into high, mid, or low categories based on first formant frequency, and grouping consonants into nasals, plosives,

fricatives, and liquids, we examined firing rates to CV pairs (Figure IV-5). While most units did not demonstrate a clear selectivity to any particular CV pair, several units showed moderate sensitivity to the combination of particular consonant and vowel classes. This may be due to a tuning of units to these particular combinations of sounds, or, because phonemes are affected by their phonemic context, this relationship may be due to changes in phoneme properties when adjacent to different types of sounds.

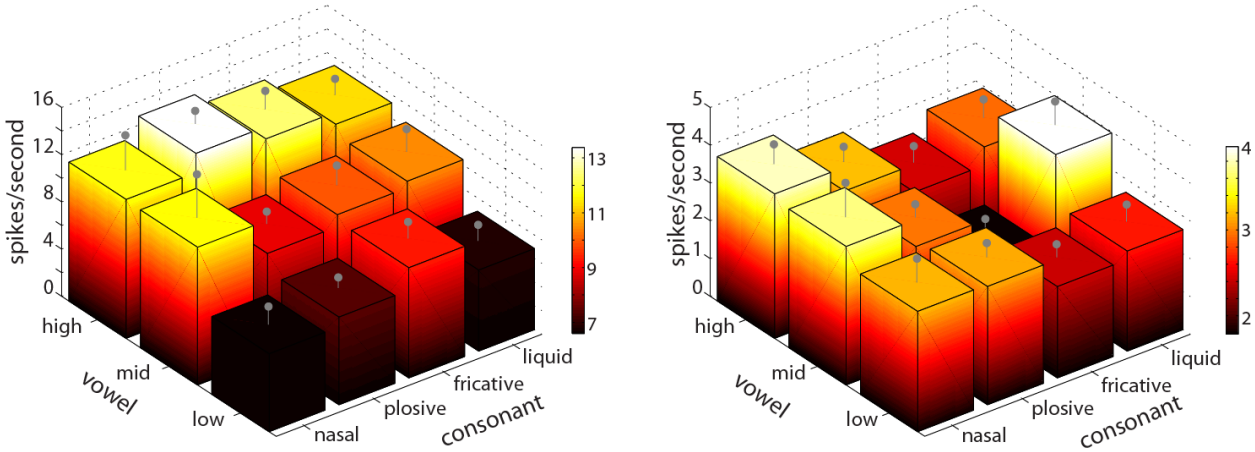

FIGURE IV-5: UNITS DEMONSTRATE MODERATE SELECTIVITY FOR CONSONANT-VOWEL PAIRS

Firing rates for unit 6a (left) and unit 24b (right) for given consonant-vowel pairs demonstrate increases in firing for high-plosive pairs (left) and high-nasal pairs (right).

To explore whether the position of phonemes within the words affected the responses of the identified units, we calculated the firing rate for each unit from 50-250ms after onset of each phoneme. Of the 22 units with significant responses to particular phonemes, 7 units responded preferentially to phonemes within the first syllable, 5 units responded preferentially to phonemes within the second syllable, and 10 responded equally well to phonemes within either syllable.

This relationship was further examined by plotting firing rate versus the latency of the phoneme after word onset (using only the phonemes the unit responds to). Units demonstrated a variety of responses including a preference to phonemes at the beginning of the word, no phoneme timing preference, or a preference to phonemes near the end of the word (Figure IV-6A). In these data, the first syllable of stimuli words was almost always stressed, meaning larger phoneme amplitudes were often found earlier in the word. Phoneme power and firing rate for these units was also correlated in a predictable fashion; if the unit preferred earlier phonemes, it also fired more to phonemes with higher power (Figure IV-6B). Thus, it is difficult to disentangle the effects of phoneme power and phoneme position in this dataset.

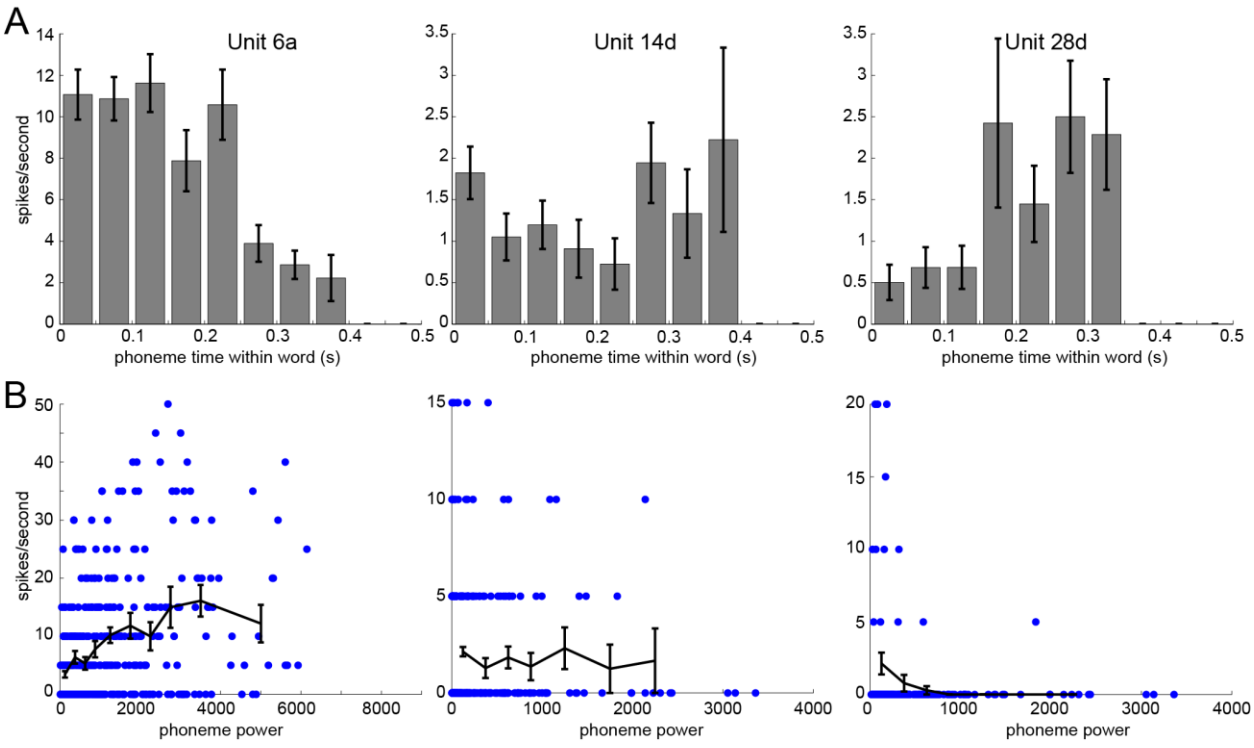

FIGURE IV-6: RESPONSES TO PHONEMES DEPEND ON AMPLITUDE AND TIMING WITHIN THE WORD.

A) These units demonstrate a difference in firing depending on the position of the preferred set of phonemes within the word (unit 6a: $\rho=-0.23$, $p<0.0001$, unit14d: $\rho=-0.001$, $p>0.05$, unit 28d: $\rho=0.19$, $p<0.0001$). B) Examples of the same three units which respond differently to varying

(Figure IV-6 continued) phoneme power (but always above perceptual threshold). The left unit shows an increasing (but saturating) firing rate to louder phonemes (ρ=0.34, $p<0.0001$), the middle unit demonstrates relative insensitivity to changing phoneme power (ρ =-0.06, $p>0.5$), and the right unit shows a decrease in firing rate depending on phoneme power (ρ =-0.15, $p<0.01$). In this figure, only preferred phonemes were used.

3.6. SPECTROTEMPORAL RECEPTIVE FIELDS

Formants and phonemes are well-established intermediate features between the acoustic signal and the word. We also attempted to characterize the units' responses in a manner less constrained by *a priori* categories by computing spectrotemporal receptive fields (STRFs) for unit 6a. The STRF computes the spectral and temporal stimulus that best drives a given unit assuming a linear relationship between firing rate and a chosen set of features. In this case, two sets of spectral features of the stimuli were utilized to compute the STRFs: power within linear frequencies from 50Hz to 4kHz, and Mel Frequency Ceptral Coefficients (MFCCs). MFCCs are a commonly used set of spectral features in automatic speech recognition and robustly represent speech information on the Mel frequency scale (which approximates the logarithmic frequency representation of the human auditory system) (Davis and Mermelstein, 1980). Furthermore, MFCCs allow for a separation of the fundamental excitation frequency of vocal chords (source) from the shape and properties of articulatory chamber (filter). In this source-filter model of speech, information that allows for discrimination of various speech-sounds is present exclusively in the "filter", while properties of the "source" are crucial in speaker-specific identification. By utilizing the first 13 Mel Frequency Cepstral Coefficients, we discard speaker-specific information that is present in the higher coefficients.

STRFs were computed using the 400 novel words from the SA task. The resulting STRF computed with linear frequency features shows a complex combination of low (50-500Hz) and high (~2.5kHz) frequency components between 0 and 100ms contributing to the firing of unit

6a (Figure IV-7A). Similarly, the STRF computed with MFCCs demonstrates a wide range of cepstral components, at a similar latency, contributing to the firing of this unit (Figure IV-7B).

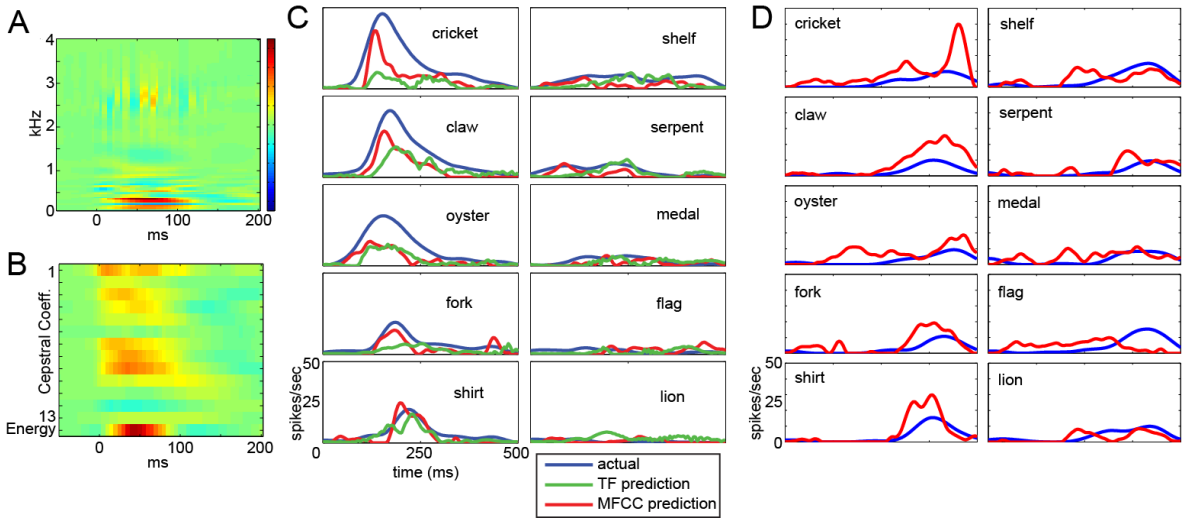

FIGURE IV-7: COMPUTED SPECTROTEMPORAL RECEPTIVE FIELDS (STRFS) CAN PREDICT UNIT FIRING RESPONSES TO WORDS.

A) STRF for unit 6a computed for linear frequency features from 50Hz to 4kHz where 0 is the firing of the unit. Power after zero indicates the frequencies that predict the cell firing at the indicated delay. B) STRF for unit 6a computed using the first 13 Mel Frequency Ceptral Coefficients (MFCCs) and an energy term. C) Predicted versus actual firing rates of unit 6a to the repeated words in the SA task. MFCC features resulted in a better prediction (R^2=0.42 vs R^2=0.16). D) Prediction of firing rates for reversed words using the STRF computed from MFCCs results in overestimation of firing rates with R^2=0.14.

To evaluate the accuracy of these representations, we attempted to predict the firing of this unit to the repeated words of the SA task by convolving the features generated from the stimuli with the computed STRFs (Figure IV-7C). The STRFs generated using the MFCCs better predicted the actual firing rates to each of the 10 repeated words than the linear frequency representation (R^2=0.42 and R^2=0.16 respectively). Despite this, both sets of features consistently underestimate the actual firing rates of this unit.

The computed MFCC STRF was used to predict the firing to the time-reversed words (Figure IV-7D). In this case, the predicted firing rates tended to overestimate the actual firing

rates, resulting in R^2=0.14. The fact that this set of acoustic features fails to adequately predict time-reversed words suggests that there are features present in words that are destroyed in time-reversal that are not being accounted for here.

3.7. RESPONSES TO WRITTEN WORDS

The patient also performed a visual word size judgment task (SV) that was equivalent to the SA task, but using written words presented on a computer screen instead of spoken words. In this task, 177 units were identified and 26% of these units demonstrated significant responses to the written words. In total, 46 units were present in both the auditory and visual tasks of which 19 did not respond to either task, 18 responded to auditory words only, and nine responded to both visual and auditory words (Figure IV-8A). All nine of the units that responded to both written and spoken words had longer response latencies to visual words than to auditory words with a mean delay of 170±31ms (mean±s.e.) (Figure IV-8C). On average, auditory words elicited an 8.04-fold increase in firing over baseline while visual words elicited a 3.02-fold increase over baseline.

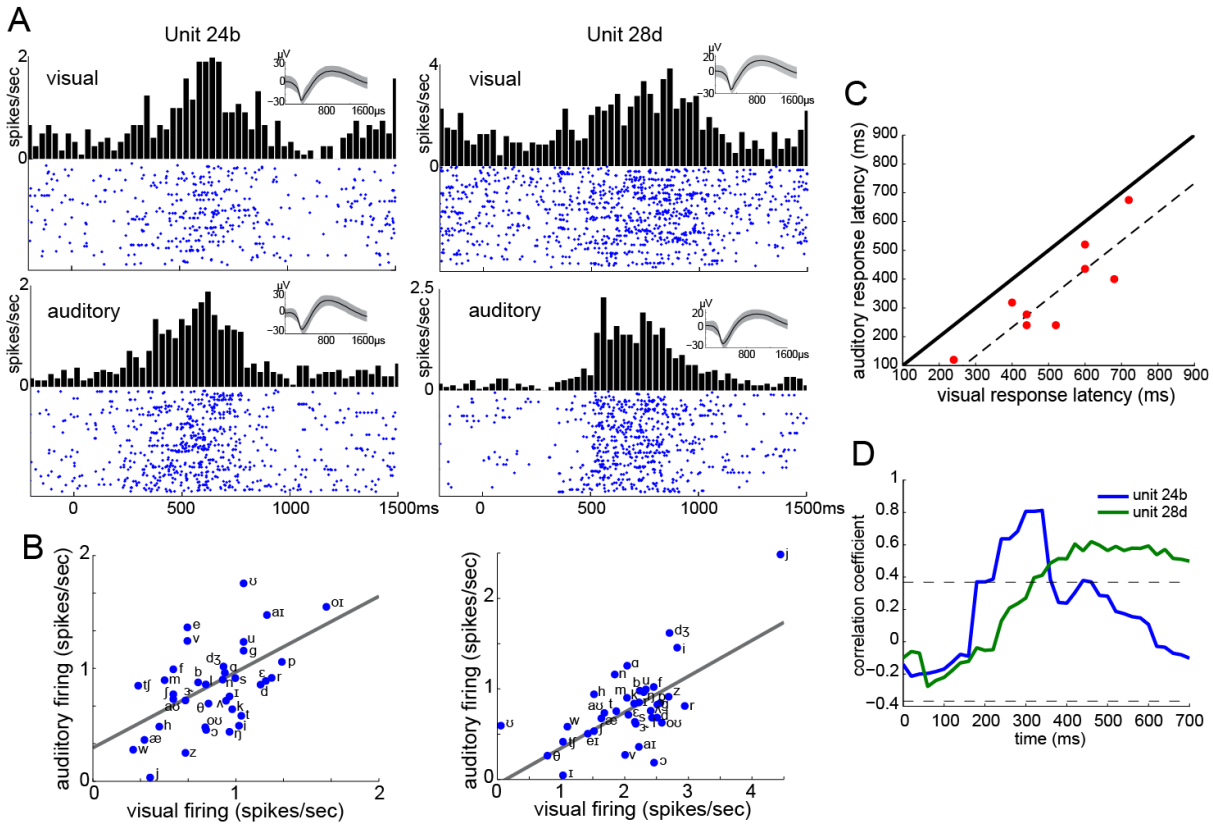

FIGURE IV-8: AUDITORY SINGLE UNITS RESPOND TO WRITTEN WORDS BASED ON PHONEMES IN PRONUNCIATION.

A) PSTHs and raster plots for two units (24b and 28d) for presentation of written (top) or spoken (bottom) words. B) Firing rates (between 0 and 1000ms) to words containing a given phoneme are significantly correlated between visual and auditory words (Spearman, Unit24b: $\rho=0.54$, Unit28d: $\rho=0.56$, $p<0.01$). Solid line represents best linear fit. C) Response latency to written words is delayed by an average of 170ms compared to auditory words (dashed-line). Solid line represents no delay between modalities. D) Correlation of firing rates between auditory words containing given phonemes between 0-1000ms and written words containing the same phonemes in 300ms sliding time windows. Windows for both visual and auditory words begin at the time post-word onset indicated on the x-axis. The onset of significant correlation for unit 24b was in the 175-475ms time window, and peak correlation was from 330-630ms. For unit 28d, significant correlation started at 325-625 and peaked from 450-750ms.

To explore whether the responses to these visual words were due to phonological properties of the words, we correlated the mean firing rates from 0-1000ms in response to phonemes present in the auditory word stimuli to the firing rates to the same phonemes in the pronunciation of the written word stimuli. The phoneme tuning properties of two of the nine units showed significant correlations with $\rho=0.54$ for unit 24b and $\rho=0.56$ for unit 28d

(Spearman, $p<0.01$) (Figure IV-8B). The peak firing rates of all but one of the remaining units were below 1 spike/sec, likely leading to poor estimates of phoneme tuning; it is possible that these other units have consistent phonological response properties to both visual and auditory words that are hidden by insufficient statistical power. Psychophysical studies have suggested that phonological recoding is stronger for low frequency words (Seidenberg, 1985). By dividing words into low and high frequency words (using the median HAL frequency of 1422), this correlation was recomputed for each set. For low frequency words, the correlation remained high at $\rho=0.44$ (unit 24b) and $\rho=0.55$ (unit 28d) (Spearman, $p<0.01$). For high frequency words, the correlation became insignificant for unit 24b ($\rho =-0.11$, $p>0.05$) and dropped for unit 28d ($\rho=0.34$, $p<0.05$).

To examine the latency of this correlation in these two units, the correlation coefficient was computed between the firing rate in response to auditory words from 0-1000ms containing given phonemes, to the firing rate in response to visual words containing the same phonemes in 300ms sliding windows starting from 0 to 700ms after stimulus onset (Figure IV-8D). Significant correlations began 175-475ms and peaked at 330-630ms for unit 24b, and began at 325-625ms and peaked at 450-750ms for unit 28d.

3.8. DIVERSITY OF UNIT TUNING ALLOWS FOR DECODING OF WORDS

In order to characterize the amount and diversity of information present in the firing rates of the identified units, we attempted to decode the 10 repeated words from the SA task using unit responses. We trained and tested a Naïve Bayes classifier after adding a single unit at each step with the goal of maximizing the accuracy (Figure IV-9A). We achieved a peak accuracy of 39.25% using 28 units (chance accuracy = 10%). Figure IV-9A demonstrates that most of the information gain can be achieved using only 10 units. The decrease in performance

after more than 30 units are added is likely due to the fact that at above 30 units, there is little information gain with each additional unit, but a continual increase in the complexity of the classifier, leading to poorer performance.

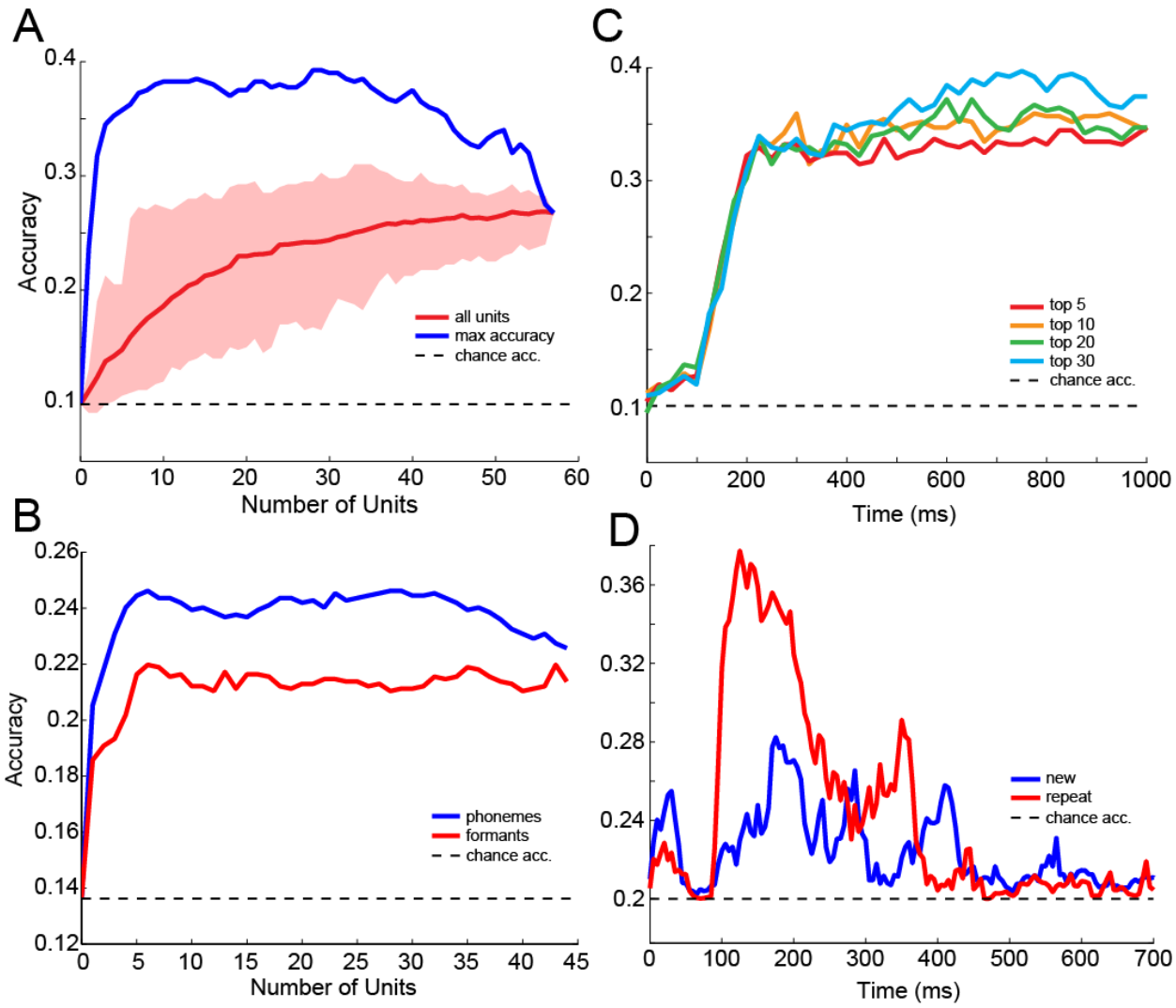

FIGURE IV-9: UNITS PROVIDE DIVERSE INFORMATION THAT ALLOWS FOR DECODING OF INDIVIDUAL WORDS AND PHONEMES.

A) Accuracy of decoding 10 words (chance = 10%) obtained by sequentially adding firing rate information from one unit at a time. The blue line is obtained by maximizing the accuracy at each step (i.e., selecting the most informative unit). The red line demonstrates the accuracy when units are randomly added. B) The decoding of either phonemes (blue) or discretized formant classes (red) demonstrate that these units provide more information on phoneme identity than formant space. C) The word decoding accuracy using a cumulative set of 25ms windows and the top 5, 10, 20, or 30 units identified by the analysis performed in panel A. Decoding is effective beginning shortly after 100ms and rises rapidly. D) First vowel decoding accuracy using a 50ms sliding

(Figure IV-9 continued) window, from the repeated words (red) and from a matched set of new words (blue). Higher performance for repeated words indicates that much more than single phoneme information is being expressed.

To explore the temporal evolution of information in this population of units, a classifier attempted to decode word information from cumulative 25ms windows starting at 0ms (Figure IV-9C). Within 200ms, 34% accuracy (chance = 10%) is reached when using the top 5 units as determined by the above analysis. Adding up to 30 units improves accuracies at longer latencies.

The same neuron adding analysis was performed for decoding vowels and yielded a peak accuracy of 24.6% (chance = 13.6%) (Figure IV-9B). To test whether location in formant space provided equivalent information, each phoneme was reclassified to the vowel that had the closest mean F_1-F_2 value in this formant space. All classes were balanced such that they had the same number of instances before and after reassignment. The accuracy of a decoder trained on this formant data yielded poorer accuracy at 22%. This suggests that these neurons encode phoneme identity better than formant values.

To test whether phonetic context affected the firing rates of the identified neurons we attempted to decode the first vowel in the repeated words in the SA task, or the first vowel in the new words. A matched set of new words were chosen so that an equivalent number of instances of each vowel were present. If these neurons only encoded information about the first vowel identity, equal decoding accuracy would be expected in both cases. The fact that the classifier was able to decode the consistent context of the repeated words better than the variable contexts found in the new words suggests that these neurons are encoding more than single phoneme identity (Figure IV-9D).

3.9. SPEAKER INVARIANCE AND SPONTANEOUS SPEECH

While it is clear that many of these units demonstrate strong responses to word-specific acoustic features, it is unclear whether they demonstrate invariance to the acoustic properties of different speakers. The speaker for the SA stimuli was male while the speaker for the WN stimuli was female. This is clearly seen in the differences in fundamental frequency of the two speakers: the fundamental frequency of the male speaker in the SA task was 113±15Hz (mean±SD) while the fundamental frequency of the female speaker in the WN task was 165±35Hz. The first two vowel formants were also different for each speaker (Figure IV-10). To check for speaker invariance, we analyzed the 37 words that were present in both the SA and WN tasks for units 6a and 6b. The spiking rates between 100 and 900ms were computed for each word, and were significantly correlated for the 37 words (Pearson, $\rho=0.38$, $p<0.05$), however, paired t-test failed to demonstrate any statistical differences ($p=0.96$, mean difference = 0.032 spikes/sec).

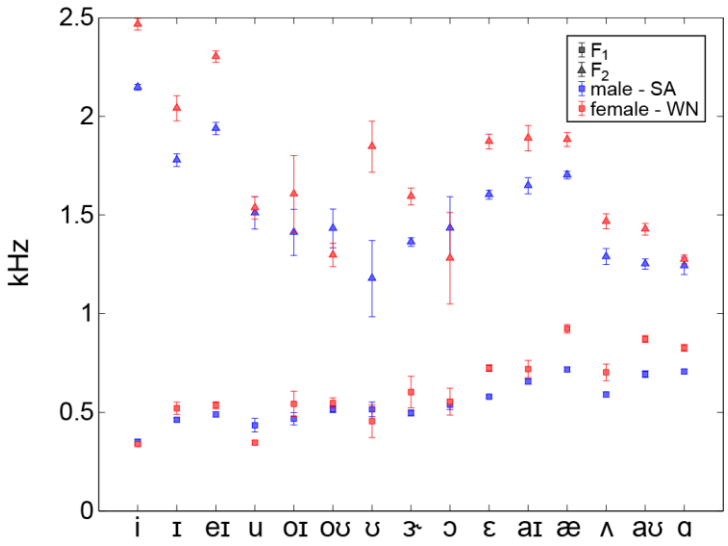

FIGURE IV-10: FORMANT DIFFERENCES FOR MALE AND FEMALE SPEAKERS IN SA AND WN TASKS.

Mean vowel first and second formant frequencies (F_1, squares, and F_2, triangles) vary between the male speaker of the SA task (blue) and the female speaker of the WN task (red). Bars indicate standard errors of the mean.

To further characterize speaker invariance, a 40 minute segment of spontaneous speech between the patient, his family, and the researchers was transcribed and all word boundaries were manually marked. When looking at the 50 most commonly produced words, the profile of firing to each of these words was significantly correlated between speakers ($\rho = 0.41$, $p<0.01$) while a paired t-test failed to indicate any significant difference in the firing rates between speakers ($p=0.35$).

3.10. SELF-VOCALIZATION AUDITORY SUPPRESSION

Previous studies have demonstrated a suppression of auditory responses to self-vocalization (Flinker et al., 2010, Baess et al., 2011, Houde et al., 2002, Heinks-Maldonado et al., 2005, Heinks-Maldonado et al., 2006, Creutzfeldt et al., 1989b). During the repetition experiment, a total of 162 units were identified, of which 42 responded to external speech. Of these 42 units, 7 showed no difference in firing to self-produced speech (Wilcoxon rank-sum,

p>0.05), 5 showed a reduced (but still significant) firing rate, and 30 showed no significant response to self-produced speech (Figure IV-11). On average, the peak firing rate to external speech was 2.43 spikes/sec higher than to self-produced speech (range = 0.21 to 13.1 spikes/sec higher). This corresponds to an average reduction to self-produced speech of 65% (range = 11% to 100% reduction).

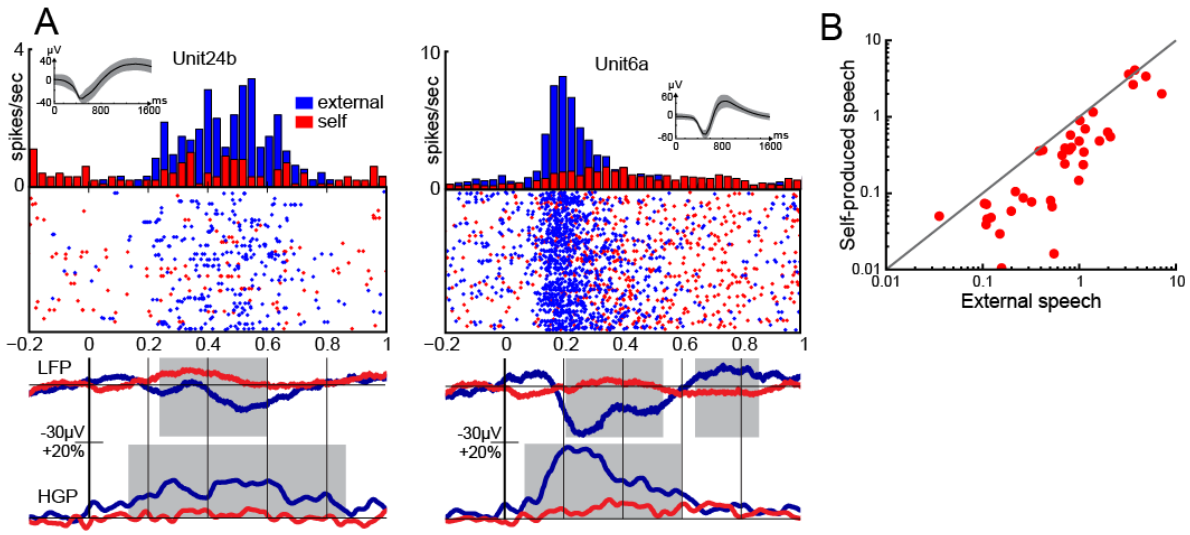

FIGURE IV-11: VARIABLE AUDITORY SUPPRESSION DURING SELF-INITIATED SPEECH PRODUCTION.

A) Representative PSTH and raster plots of auditory responses during self-initiated speech (red) versus external speech (blue) during the repetition task for an excitatory cell (left) and inhibitory cell (right). Below, the averaged LFP and high gamma power (70-100Hz, bottom) for the corresponding electrodes demonstrate minimal changes to self-produced speech suggesting suppression in earlier areas. B) Firing rates of responding units during self-produced speech versus external speech. The unity line indicates no change in firing rate between the two conditions with points lying below the line indicating a reduction in firing during self-produced speech.

Five putative inhibitory interneurons were present within the 42 units that responded to external speech. All five units demonstrated a decrease in firing rate to self-produced speech relative to external speech (2.36 spikes/sec vs. 1.32 spikes/sec). Additionally, averaged local field potential (LFP) of the corresponding electrodes demonstrate minimal responses to self-

produced speech while showing large responses to external speech at all latencies (Figure IV-11A).

4. DISCUSSION

In this chapter, we characterized single units recorded from left anterior superior temporal gyrus that demonstrated highly selective responses to spoken words with little or no responses to pure-tones, environmental sounds, or vocoder-transformed speech. Many cells fired specifically to particular words or subsets of phonemes, demonstrated spatial-organization of tuning properties, and were suppressed in response to self-produced speech. Computed spectrotemporal receptive fields predicted responses to spoken words and a period of spontaneous conversation demonstrated invariance to speaker. Some units showed correlated responses to phonemic properties of visual and auditory words.

The classical model of the auditory language processing stream involves the posterior, rather than anterior, superior temporal gyrus; speech information is posited to initially activate primary auditory cortex in Heschel's gyrus, and then move back towards Wernicke's area as processing becomes more complex (Boatman et al., 1995, Chang et al., 2010, Crone et al., 2001a, Desai et al., 2008, Geschwind and Levitsky, 1968, Wernicke, 1874, Arnott et al., 2004). The existence of single units in aSTG that are tuned to speech sounds suggests that this processing may be more distributed than initially thought. Instead, processing of speech for semantic understanding may proceed in an anterior-ventral direction, as suggested by the dual-pathway model for speech processing wherein the aSTG is an early stage in the auditory speech-specific "what" stream (Rauschecker and Scott, 2009, Hickok and Poeppel, 2007, Saur et al., 2008). In this view, the neurons reported in this chapter are likely analogous to inferotemporal (IT) cells within the ventral visual stream that respond to increasingly complex

visual objects (Tanaka, 1996). Similar to these IT visual cells, the units described here also demonstrate spatial organization of selectivity (Fujita et al., 1992). The "what" stream likely continues to anteroventral temporal neurons that fire differentially to words referring to objects versus animals (Chan et al., 2011a), and even to particular words (Heit et al., 1988). Thus, although highly and selectively responsive to spoken words, firing by aSTG neurons do not reflect semantic modulations. Furthermore, macroelectrode electrocorticography (ECoG) recordings directly over the microelectrode site fail to show language-specific responses, suggesting that this activity may be very focal, engaging a limited neuronal population (Figure IV-12). Despite this, posterior temporal areas do show robust ECoG responses, and this may explain why speech-specific activation of aSTG has been relatively unseen when compared to pSTG activity in recording modalities with lower spatial resolution.

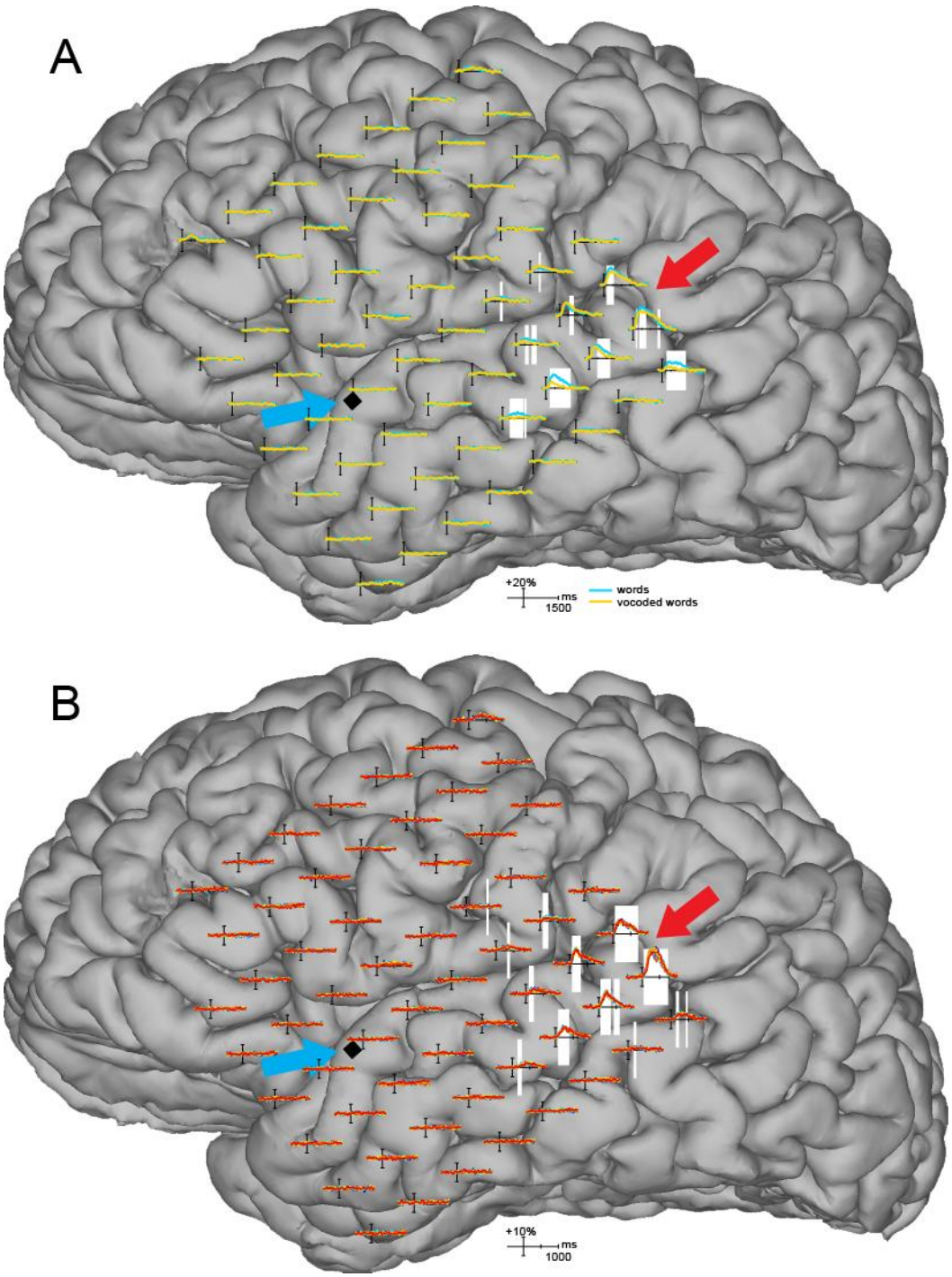

FIGURE IV-12: MACROELECTRODE GRID HIGH-GAMMA POWER FAILS TO SHOW WORD-SPECIFIC RESPONSES NEAR THE MICROELECTRODE ARRAY.

High gamma power from 70-200Hz was computed over the macroelectrode grid for the words versus noise-vocoded words in the WN task (A) and for the 10 repeated words in the SA task (B). The largest responses are seen around posterior superior temporal lobe near Wernicke's area (red arrows). The white shading indicates areas that demonstrate significant differences between words and noise, or the 10 repeated words. Macroelectrode contacts near the microelectrode array

(Figure IV-12 continued) (blue arrows) fail to demonstrate high gamma power responses, despite the robust speech-specific unit firing directly under these electrodes.

The cells reported here demonstrate significant responses to specific subsets of phonemes, and in the most robust cell, to a subset of vowel sounds. The observation that unit firing is invariant to speaker, regardless of differences in formant frequencies, F_0, F_1, and F_2 underlying the production of these vowel sounds, suggests that these units are tuned in vowel-space rather than formant space. Furthermore, the decoding of phonemes resulted in significantly higher accuracies than the decoding of F_1, F_2-derived classes.

The results from the STRF analysis provide additional evidence that low-level acoustic features fail to fully characterize the response properties of these units. Power in frequency bands is a poor predictor of firing rate when compared to MFCCs that model the phonetic and discard speaker-specific information in speech. However, even the MFCCs fail to robustly predict the firing rate in time-reversed speech. This may suggest that MFCCs do not completely capture some other high-level acoustic features of words, and that time-reversal destroys, or at least reduces the presence of, these features. Phoneme identity is one such feature; time reversing the acoustic waveform associated with a phoneme, especially consonants, often produces a percept that is non-phonemic in nature. It is possible that these units are tuned to high-level properties of phonemes that we are unable to completely characterize.

Response latencies to words varied widely between the identified units, and in fact, different units demonstrated different levels of firing depending on the timing within the word at which the preferred phonemes occur. Because most of the bisyllabic stimulus words were stressed on the first syllable, it is difficult to disentangle the effects of phoneme position within a word, and phoneme stress. Despite this, one might imagine that a population of units which

respond to different sets of phonemes in different contexts/timings could serve as the building blocks by which any word may be represented.

The current results also may be relevant to the long-standing controversy as to whether written words must activate their phonological representations in order to be understood. One theory posits that written words have a direct access to meaning as well as an indirect access via phonological recoding (Seidenberg, 1985, Yvert et al., 2012, Coltheart et al., 1993). This theory suggests that skilled readers reading high frequency words access lexical representation too quickly for phonological information to contribute. In contrast, another model suggests that written words necessarily undergo phonological processing before lexical identification, regardless of word frequency or task demands (Frost, 1998). Several studies have compared scalp ERPs to written words, pseudohomophones (nonwords that sound like an actual word), and other control nonwords (Braun et al., 2009, Newman and Connolly, 2004). The differential response to pseudohomophones is taken to represent a conflict between orthographic and phonological information at the lexical level, and therefore the presence of ERP differences prior to 200ms has been taken as evidence for obligatory early phonological recoding of written words (Braun et al., 2009). Despite this, the observed effects are quite small, several studies have failed to find them (Ziegler et al., 1999, Newman and Connolly, 2004) or estimate their latency as ~300ms (Niznikiewicz and Squires, 1996), and their localization is unclear. The visual word form area in the fusiform gyrus produces its major response at this latency, and it is highly sensitive to the combinatorial frequency of letters (Binder et al., 2006, Vinckier et al., 2007), raising the possibility that some signal could arise from this area, perhaps analogous to the automatic mismatch negativity commonly studied in the auditory modality (Näätänen and Picton, 1987).

While MEG constrained with MRI and fMRI has localized activity in the posterior superior temporal cortex to visual words beginning at around 200ms (Marinkovic, 2004, Dale et al., 2000), there is no evidence that this activity represents phonological recoding. Conversely, although fMRI identifies this general region as activated by written words in tasks that require phonological recoding (Fiebach et al., 2002), it is not possible to know if such activation is before or after lexical access. In contrast to these previous studies, the finding presented here is a direct demonstration of phonological recoding: unit firing that is correlated between spoken auditory phonemes and phonemes present in the idealized pronunciation of visual words. Lexico-semantic access for visual words is generally thought to occur by ~240ms (Halgren et al., 2002, Kutas and Federmeier, 2011). In our data, the firing of one cell reflecting phonological recoding of written words began slightly before this time in the window beginning 175ms after word onset. Furthermore, high-frequency words demonstrated reduced correlation between phonemes in visual and auditory words, presumably reflecting a smaller need for phonological recoding. Thus, these limited data are consistent with the dual-route hypothesis of phonological recoding, in that we demonstrate that neural activity with the expected characteristics occurs in the STG at a latency that may allow it to contribute to word identification. However, it is possible that phonological recoding is not evoked by all words, since the auditory-visual correlation was greatly decreased for high frequency words.

The use of the microelectrode array in this chapter allowed for the examination of spatial organization that previous studies have been unable to explore. Interestingly, we found that nearby cells often had correlated response properties, but this correlation disappears at distances of over 1mm. This may suggest that even in high-order processing areas that do not have a clear spatial or spectral "space" (such as orientation or frequency), nearby cortical columns perform similar processing tasks. It is also important to note that more general

response characteristics (e.g. whether the unit responded to auditory word stimuli at all) showed broader spatial correlation than more specific response characteristics (e.g. 10-word response profile), which tended to exhibit narrower spatial correlations. This is similar to the columnar organization of inferotemporal visual cells that demonstrate spatial organization of visual object selectivity (Fujita et al., 1992, Tanaka, 1996). The consistency of firing profiles across the set of words for different units recorded at a given contact could be reflected in the population activity (high gamma power) recorded by the same contact.

Studies using PET, EEG, MEG and intraoperative microelectrodes have shown that auditory cortex is suppressed during self-produced speech, when compared to external speech perception, and it has been suggested that this is a result of speech-feedback monitoring (Baess et al., 2011, Heinks-Maldonado et al., 2005, Heinks-Maldonado et al., 2006, Houde et al., 2002, Tourville et al., 2008, Christoffels et al., 2007, Curio et al., 2000, Numminen et al., 1999, Paus et al., 1996, Creutzfeldt et al., 1989b). These studies have suggested that this phenomenon occurs globally across auditory cortex, however units in primary auditory cortex of primates (Eliades and Wang, 2005) have demonstrated a diversity of responses to self-produced vocalizations. Flinker et al. (2010) have shown that ECoG recordings in humans demonstrate a varying degree of suppression across regions of auditory cortex. In this chapter, we show that this variability in is present to an even smaller scale; single units within a $4mm^2$ area demonstrate variable amounts of suppression by self-produced speech. Our additional finding that putative inhibitory interneurons also exhibit reduced firing, and that LFPs to self-produced speech are suppressed from their onset, suggest that the suppression begins at an earlier processing stage and that decreased local firing is due to decreased input.

It is important to note that these recordings come from the unique case of a single patient with epilepsy. The cortical location containing the microelectrode was included in the final resection, and subsequent staining and histology failed to find abnormal pathology at the array site. Furthermore, the patient's seizures were found to start in medial temporal sites, making it less likely that aSTG was actively involved in seizure initiation. However, we cannot rule out the possibility that medications or long-standing epilepsy affected the responses we recorded, or that other idiosyncrasies in this patient's data exist that we could not assess. It is also possible that, despite the large number of tasks this patient performed, we failed to present a critical stimulus category.

Taken together, these data suggest that the anterior STG contains a spatially-organized processing unit specialized for extracting lexical identity from acoustic stimuli, lying midway between acoustic input in medial Heschl's gyrus and supramodal semantic representations in anteroventral temporal cortex. This module encodes high order acoustic/phonetic information during the perception of both spoken and written words, suggesting the aSTG is involved in phonological recoding during reading. Single units robustly represent perceptual phonemic information, and it is possible that a small population of cells, each encoding a different set of phonemes in different phonological contexts, could represent the acoustic form of a specific word.

V. CONCLUSIONS

In this dissertation, we began by examining semantic representations from extracranial electro/magnetophysiological recordings. We demonstrated that machine learning techniques could robustly decode both category and word-specific information from both EEG and MEG, and that this information was highly distributed across the cortical surface. Furthermore, information was consistent across subjects as well as between presentation modalities. Interestingly, the ability to decode individual words surpassed the ability to decode the two semantic categories. Several possibilities may account for this difference. It is possible that animal/object category-specific information is localized to an area that is not well recorded through extracranial techniques, or that the features utilized for the decoding did not adequately capture this information. Also, if semantic concepts are represented by the combination of features which define an object, then discriminating between individual words would potentially use many of these features to distinguish between the two concepts. However, the living/nonliving contrast is only one such contrast, potentially providing fewer distinguishing features and making it more difficult to successfully decode these two semantic categories.

We subsequently proceeded to intracranial recordings that provide higher spatial resolution, and allow for the examination of structures that are more difficult to record from in EEG and MEG. By using depth electrodes to examine differences in local field potential and gamma band power, we observed robust differences between animals and nonliving objects in anterior collateral sulcus and occipito-temporal sulcus. These differences began at a latency of approximately 350-400ms and were observed in both the visual and auditory versions of the language task, suggesting a semantic basis for this activity. The use of laminar microelectrodes allowed for the study of medial temporal structures and showed that the category-selectivity

was present in entorhinal cortex, perirhinal cortex, and inferotemporal cortex, and came with the first pass of activity through these areas beginning at approximately 200ms. Multi-unit activity and CSD measures provided features by which accurate decoding of semantic category was possible. A small number of perirhinal single units were also observed that demonstrated category-specific differences in firing rate. These observations suggest that the avTL is important for semantic representations, and that previously seen posterior temporal-occipital animal/object differences are likely differences in structural/visual, rather than conceptual, representations.

Finally, we examined the transformation of auditory signals into word-specific representations at an even finer spatial scale. A unique opportunity was described in which a microelectrode array was implanted in superior anterior temporal lobe that allowed for the recording of a large number of auditory-responsive neurons in a patient with epilepsy. These units demonstrated robust firing to particular speech sounds and are insensitive to equally complex non-speech sounds. These units also fired to written words in a manner consistent with the phonological properties underlying the pronunciation of the word, providing direct neural evidence of the phonological recoding hypothesis in reading. Furthermore, the firing rates of a small number of units allowed for the decoding of individual word identity. These results suggest that the superior anterior temporal lobe supports population coding of the acoustic/phonetic features of speech and that these neurons may be the building blocks by which auditory word-specific representations are built.

In this final chapter, I will synthesize the results presented in the previous three chapters and place them in the context of the whole of the language processing system. I shall also discuss common themes and pathways observed throughout this dissertation and present

an overall model of the processing of visual and auditory words. I will then discuss the potential applications of decoding language information and explore possible ways the accuracy of such decoders might be improved. Finally, I will comment on potential future directions and natural extensions of the work presented here.

1. MODEL OF WORD PROCESSING

1.1. HIERARCHICAL PROCESSING AND GRANDMOTHER CELLS IN LANGUAGE SYSTEM

The results presented in Chapter IV demonstrate that like the visual and auditory systems, word recognition likely proceeds in a hierarchical fashion. For spoken words, this would begin with pure frequency selectivity in primary auditory cortex from which more complex time-frequency features would be built, leading to speech-specific acoustic features and eventually auditory word-specific representations. For written words, this would proceed through V1 from which simple edges and shapes would be derived, leading to representations of individual graphemes and eventually full visual words. While in both these cases, there is clear evidence that low level stimuli (e.g. edges of light intensity or simple frequency tones) are encoded at the single unit level (Hubel and Wiesel, 1962, Bitterman et al., 2008, Howard et al., 1996), it is still unclear how complex or abstract a representation might be that a single neuron could be tuned to. One might imagine that after a certain point, highly complex stimuli, like full words, are represented only by the population activity of many neurons.

In the visual system, one study by Mormann et al. (2008) suggests that this hierarchical processing applies to representation of complex visual objects at the level of single units as far as the medial temporal lobe. Likewise, it is also possible that as we proceed upstream, single neurons continue to be tuned to more and more complex language representations. In the

visual system, the idea that an individual cell would be tuned to a highly complex and very specific visual object is known as the Grandmother Cell hypothesis (Desimone, 1991). This hypothesis posits that the visual representation of one's grandmother might be represented by a single neuron. In an extension of this for the representation of speech information, one might imagine that individual cells are tuned to particular auditory word representations, and this might lead to single cells that are tuned to supramodal word representations that may lead to cells that encode specific supramodal semantic concepts.

While the Grandmother Cell hypothesis is difficult to test, Chapter III demonstrated that single cells in avTL demonstrate category-selectivity. Furthermore, Heit et al. (1988) have demonstrated single unit firing in human medial temporal lobe to specific visual words, Kreiman et al. (2000) have demonstrated visual category selective cells in the same region, and Quiroga et al. (2005) have shown visual object selective cells in MTL. While none of these studies suggest that these cells only respond to the particular stimulus they responded best to (e.g. pictures of Jennifer Aniston), they do claim that these data strongly suggest that the representations are sparse, i.e. that any given cell responds to a very limited set of complex stimuli (Quiroga et al., 2008). Furthermore, it is highly unlikely that the recorded cells in these studies are the only cells that are selectively tuned for these complex objects. In fact, Waydo et al. (2006) suggests that on the order of two million cells represent a given percept in the visual system. Unfortunately, it is impossible to ever fully test the infinite space of all possible stimuli to determine how specific the tuning of such cells is, but the Grandmother Cell hypothesis is likely an overly extreme position.

Quiroga et al. (2009) have called these neurons in human MTL "concept cells" because they often respond to multimodal stimuli referring to the same "concept". For example, a cell

that fires to an image of Hale Berry may also fire to the written and spoken name of the actress. While it is possible that these cells encode very specific high-level amodal conceptual knowledge, it is also possible that they instead encode high level visual information that is activated when encountering a word that refers to this concept. Regardless of what these single units represent, the data in this dissertation demonstrate that, at the very least, semantic/conceptual information is robustly represented in these medial temporal areas by small populations of neurons (Section III.3.3).

While the visual system clearly demonstrates single unit hierarchical processing for highly complex stimuli, whether this is also true for the evolutionarily younger development of words and language is unclear. The results in Chapter IV demonstrate that in the auditory word processing stream, fairly high level speech sounds are still encoded by single neurons. The lateral temporal cells observed here are clearly along the hierarchical stream of speech processing towards representation of auditory words. While the data in this dissertation is unable to answer whether single units also represent full word (auditory or visual) lexical representations, this possibility exists. There are two potential models for the interface between the units shown here, and the semantic/conceptual representation associated with words (Figure V-1). First, there may exist single units that sparsely represent a full auditory word/lexical item and receive converging information from the high-level acoustic-phonetic neurons observed here. The report by Heit et al. (1988) provides some evidence for this possibility. These neurons may subsequently activate a diverse population of cells (potentially located in MTL and other cortical areas) that represent the full conceptual knowledge associated with that lexical item (Figure V-1A). The possibility also exists that there are no single neurons that represent these full acoustic words and instead the representation of a lexical item is solely determined by the population activity of the underlying phonetically-

tuned cells. This distributed representation may then go on to activate the distributed semantic representation of the lexical item (Figure V-1B). While the testing of these models is difficult, performing additional single unit recordings in left anterior temporal lobe may provide informative data in this regard.

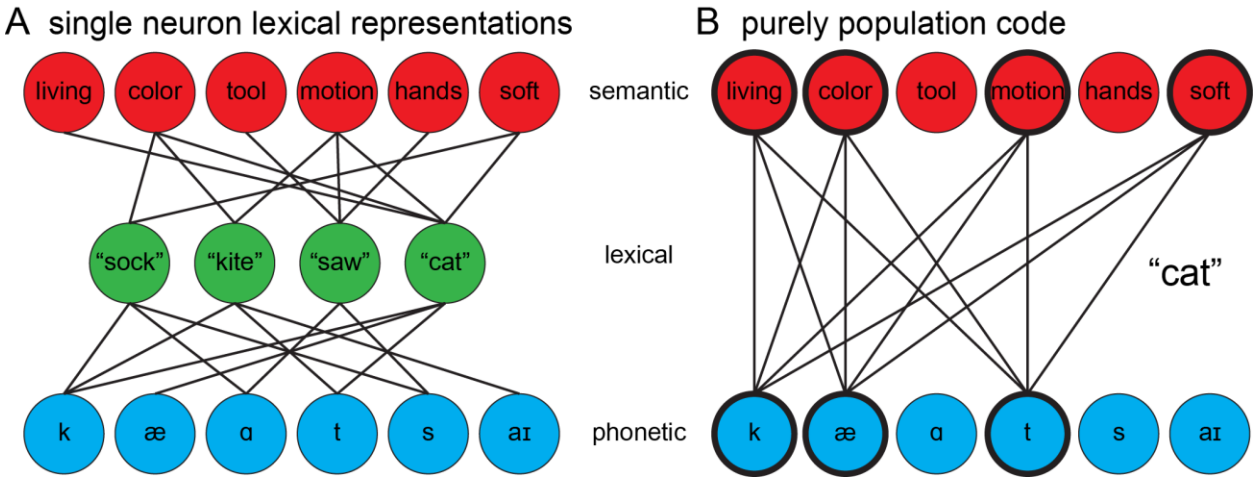

FIGURE V-1: TWO MODELS OF THE INTERFACE BETWEEN PHONETIC AND CONCEPTUAL INFORMATION

In each of these models, circles represent individual populations of single neurons. A) In this model, single neurons represent the representation of a spoken word or lexical item by combining the information from individual high-level acoustic/phonetic cells. These single neurons can then activate the potentially distributed set of cells that represent the various aspects of the conceptual knowledge regarding this lexical item. B) The alternative is that a lexical item is purely represented by the population activity of the phonetic cells and/or semantic/conceptual cells and there are no single units that are solely responsible for the representation of a full word acoustic form.

The processing of written words proceeds along the ventral visual stream and studies have suggested that this stream is also structured hierarchically (Vinckier et al., 2007, Nobre et al., 1994, Cohen et al., 2000, Cohen et al., 2002, Dehaene et al., 2005, Dehaene et al., 2002, Vinckier et al., 2006). In this case, letter-like symbols are first constructed from lines and edges of particular orientations and curvature, from which prelexical bigrams and larger combinations may be represented, finally yielding representations of meaningful words (Dehaene et al., 2002, Vinckier et al., 2007). Similar to the speaker invariance seen in the single

unit recordings presented in Chapter IV, the visual word form area (VWFA) demonstrates invariance to the case (upper or lowercase) of text as well as the font (Dehaene et al., 2001). While this dissertation has demonstrated that auditory neurons can be tuned specifically to the acoustic-phonetic components of spoken words, it is unknown whether single units along the visual word processing pathway are tuned to each of these written word elements. Written language is even younger, evolutionarily, than spoken language, and this may make it more likely that purely population coding is present for written language representations.

1.2. SPECIALIZATION AND SEPARATION OF LANGUAGE PROCESSING

As mentioned previously, language is a relatively new evolutionary development when compared to general audition, object recognition, or other sensory-motor functions. It is therefore natural to question whether the processing of speech and written words occurs within areas and circuits that originally process, and potentially continue to process, other types of auditory or visual stimuli. In other words, at what point, if ever, do language processing streams diverge from general auditory/visual processing into specialized areas? This question is highly contentious and numerous debates have occurred on various aspects of this subject (Cohen and Dehaene, 2004, Liberman and Mattingly, 1989, Price et al., 2005, Price and Devlin, 2003, Price et al., 2003b). Cohen and Dehaene (2004), described three potential forms of specialization when describing the VWFA: functional specialization, reproducible localization, and regional selectivity. Functional specialization is the idea that a particular neural system is adapted for performing a specific functional task, regardless of how this system is localized. Reproducible localization refers to the idea that the processing in question is performed in a localized brain area that is consistent between subjects. Finally, regional

selectivity is the hypothesis that regions of cortex are devoted exclusively to the specific functional task.

In terms of the speech recognition, several studies suggest that the processing of speech occurs, at least at some level, in dedicated neuronal circuitry (Vouloumanos et al., 2001, Turkeltaub and Coslett, 2010, Obleser et al., 2006, Obleser et al., 2010, Scott et al., 2000, Liebenthal et al., 2005, Whalen et al., 2006), while other have rejected the idea that speech processing occurs in specialized areas, but instead occur in areas that also support non-verbal sound processing, pitch monitoring, or conceptual processing (Price et al., 2005, Price et al., 2003b). At the very least, many of these studies have demonstrated that speech stimuli, when compared to non-linguistic auditory stimuli, elicit consistent activity in STG and STS, largely on the left side, demonstrating reproducible localization (Vouloumanos et al., 2001, Whalen et al., 2006, Liebenthal et al., 2005, Binder et al., 1996). Whalen et al. (2006) even hypothesizes that speech specialization, in the functional sense, begins as early as primary auditory cortex in Heschl's Gyrus.

The results presented in Chapter IV are consistent with the localization demonstrated in these earlier studies, and the fact that these neurons robustly respond to speech stimuli but largely fail to respond to other non-speech stimuli (e.g. vocoder transformed speech, pure tones, environmental sounds) suggests that this set of neurons demonstrates functional specialization and regional selectivity. The data are in line with the idea that, at least at the level of phonetic information, a specialized speech processing pathway exists in anterior superior temporal lobe that is not utilized for processing general nonspeech stimuli. Despite this, it is nearly impossible to conclusively prove specialization due to the infinite space of stimuli that cannot be exhaustively searched in any experimental paradigm.

While specialization for the processing of speech, which is tens to hundreds of thousands of years old, may plausibly be an evolutionarily-derived separate pathway, the development of written language (approx. five thousand years old) is too young to have evolutionary origins (Tallerman and Gibson, 2012). On an evolutionary time scale, written language is a very young invention, and until relatively recently, only a small portion of the human population was able to read. Despite this, several authors argue that a specialized system exists for the explicit purpose of reading (Cohen and Dehaene, 2004, Cohen et al., 2000, Dehaene et al., 2005, Dehaene et al., 2002, Dehaene et al., 2001). This specialized function, therefore, must be adapted from existing neural architecture that may have originally been designed for other visual processing tasks. There is no *a priori* reason that the localization of this function should be consistent between subjects given the plasticity of the brain. Dehaene and Cohen (2007) have put forth the Neuronal Recycling hypothesis in which they state that the brain has a set of architectures, and corresponding constraints, for the processing of different evolutionarily-derived functions. They suggest that culturally-acquired functions such as reading and arithmetic, in a sense, find a "foster home" in a set of neural circuits that are sufficiently close in processing to allow for a repurposing of this segment of cortex. In the case of reading, they argue that this position (along the mesial edge of the posterior occipito-temporal sulcus) in the ventral visual stream is optimized for the detection of visual objects of the correct complexity and form, and this originally general visual area has been repurposed for the representation of orthographic information (Dehaene and Cohen, 2007).

It is unclear whether such "neuronal recycling" is at play in the speech processing system. It is possible that specific human evolution has provided a distinct pathway for speech, however it is also possible that speech recognition is a repurposing of the general hierarchical stream of auditory processing. Linguists and psychologists have noted that a "critical period"

exists for language acquisition, after which it becomes nearly impossible to acquire a first language, and becomes increasingly difficult to fluently acquire a second language (Johnson and Newport, 1989, Lenneberg, 1967). At the very least, this suggests that the acquisition of speech processing partially depends on the early plasticity of the human brain.

1.3. PHONOLOGICAL BASIS OF LANGUAGE SYSTEM

The evidence for the phonological recoding of visual word stimuli presented in Section IV.3.7 raises questions regarding the relationship between orthographic and phonological representations. Given the preceding discussion regarding the relative age of spoken versus written language, it may be possible that the recognition of written words is inextricably linked to the speech recognition processing pathway. One extreme position might suggest that the mental lexicon exists purely in a phonological space, and that no direct pathway exists from sublexical orthographic elements to lexical word representations and conceptual knowledge. This would require written words to be necessarily recoded from orthographic elements to an acoustic-phonetic representation before the corresponding semantic knowledge of the lexical item can be accessed (Frost, 1998). In other words, the development of written language, while potentially requiring the neuronal recycling of visual cortex, is built such that it feeds orthographic processing into the phonological processing system rather than developing a separate visual lexicon.

While some claim that a purely phonological route for reading may be too strong due to the presence of pre-linguistically deaf individuals who are still able to read (Seidenberg, 1985), others have shown that these individuals still demonstrate sensitivity to the phonological properties of written words (Dodd and Hermelin, 1977, Hanson and Fowler, 1987). A weaker hypothesis is that while phonological recoding of a written word may occur when a task

requires it, it is not necessary for lexical access, and may in fact occur after lexical/semantic processing (Humphreys et al., 1982). In this view, phonological processing of visual words may be due to feedback from lexical processing areas rather than a bottom up translation of orthography to phonology. This model would also imply that the latency of phonological access would occur later than lexical access (Braun et al., 2009). A number of studies have attempted to examine the latency of phonological access during written word recognition, however results remain highly variable and do not provide clear conclusions in this regard (Braun et al., 2009, Ziegler et al., 1999, Barnea and Breznitz, 1998, Bentin et al., 1999, Newman and Connolly, 2004, Niznikiewicz and Squires, 1996, Proverbio et al., 2004).

The third possibility, known as the dual route hypothesis, is a combination of the two options discussed above that suggests that visual words can take one of two paths to access lexical information (Humphreys and Evett, 1985, Coltheart et al., 1993, Seidenberg, 1985). Visual information may either directly activate the corresponding lexical item, or a translation of orthography to phonology may lead to a mediated recognition via the phonological path. Some versions of this hypothesis suggest that for common words, phonological information is unnecessary, but for infrequently encountered or novel words, the phonological route is utilized (Seidenberg, 1985). The results presented in Section IV.3.7 suggest this aspect of the hypothesis may be true, however, it is also clear the phonological recoding occurs even in a task that does not explicitly require it. Further experiments would need to be done to explore whether this information is accessed pre or post-lexically by determining the relative latencies of orthographic encoding, lexical access, and phonological recoding. Examination of simultaneous intracranial macroelectrode recordings of anterior STG, anterior fusiform, and posterior fusiform gyrus may allow for determination of the sequence of processing when written words are used as stimuli.

1.4. SEMANTIC HUB REVISITED

The end goal of the word recognition system is to interface with the semantic representation of the lexical item. Chapter II very clearly demonstrated that semantic representations are highly distributed over the cortical surface. While it is still debated whether a semantic hub is necessary for the coordination of this semantic knowledge, the results in Chapter III, at the very least, suggest that the anterior temporal lobe is important for semantic representations. One possibility is that this area of the avTL is in fact part of the semantic hub system, and the semantic category differences observed in Chapter III are a manifestation of the differences in retrieval/coordination of conceptual knowledge needed to represent animals or manmade objects. This part of the avTL may not directly encode semantic features of animals or objects, but may instead be responsible for retrieving that information from distributed cortical areas, and this difference in feature retrieval is captured in the LFP/gamma-band/MUA differences observed in our intracranial recordings. The available data is, unfortunately, insufficient to determine whether this might be the case.

If the anterior temporal lobe is the location of a semantic hub, it should be amodal and demonstrate clear distinctions between many groups of semantic categories. Given sufficient recording coverage of the aTL, we would expect category-specific activation in the semantic hub to be identical regardless of input or output modality. In addition, we would expect to be able to decode category information from a number of different conceptual classes (e.g. animals, tools, dwellings, fruit, faces) solely from information obtained from the aTL. Therefore, discriminating between an orange and a basketball may be difficult when using primarily visual information, but this might be an easy task in the aTL. Similarly, the reverse might be true when discriminating between a clam and a fish due to mutual classification as "seafood" despite highly distinct visual properties.

When viewing the hypothesis of a semantic hub in terms of the theories of semantic organization presented in Section I.1.5, it would seem plausible that in actuality, a combination of the "feature-based" and "category-driven" theories could be in effect (Patterson et al., 2007, Caramazza and Shelton, 1998). The distributed network of potentially modality-specific information may embody the features that define any given semantic concept, but this information may be coordinated and organized in a category-specific fashion by the semantic hub in the ATL. While the importance of the anterior temporal lobe in semantic representations is slowly becoming more recognized, definitively demonstrating that it has all the characteristics of an amodal region for the coordination of semantic knowledge is far from complete.

1.5. OVERALL MODEL OF WORD RECOGNITION

We can synthesize the results in this dissertation and previous literature into an overall model of word processing (Figure V-2). Incoming words initially enter either through the primary visual or auditory cortices. In the case of speech, after low-level acoustic processing, information is split into two streams (Saur et al., 2008, Hickok and Poeppel, 2007). The dorsal stream is responsible for the mapping of sound onto articulation or spatial location and proceeds posteriorly from Heschl's Gyrus along the STG to Wernicke's area, and may eventually lead to Broca's area in the inferior frontal lobe, or other prefrontal areas. The anterior/ventral stream is responsible for mapping sound to meaning and proceeds forward and inferiorly from Heschl's to aSTG and STS and eventually interfaces with lexical and semantic representations, presumably located in anterior-inferior temporal areas. This dual pathway theory of speech perception fits nicely with the observed single unit tuning for discriminative speech sounds in aSTG as well as the observed phonemic specificity other studies have found in pSTG (Crone et

al., 2001a, Desai et al., 2008, Steinschneider et al., 2011, Turkeltaub and Coslett, 2010). In both cases, phonological information must be extracted, but for different purposes. In the anterior/ventral "what" stream, we would expect to find representations of lexical identity, while the posterior/dorsal pathway would not necessarily contain this information. We would also expect to encounter representations of phonemes that could be grouped in terms of place (e.g. bilabial, labio-dental, palatal) and manner of articulation (e.g. frication, rounding, voicing) in this dorsal stream. Given the presence of single units that represent speech sounds for lexical identification in aSTG, we might also predict that these phonetic-articulatory representations are represented by single units within the pSTG or elsewhere in the dorsal pathway. The idea of the dorsal stream being important for recognizing articulatory features is also consistent with lesions of this area leading to cases of conduction aphasia in which a patient is unable to repeat a spoken word (Damasio and Damasio, 1980).

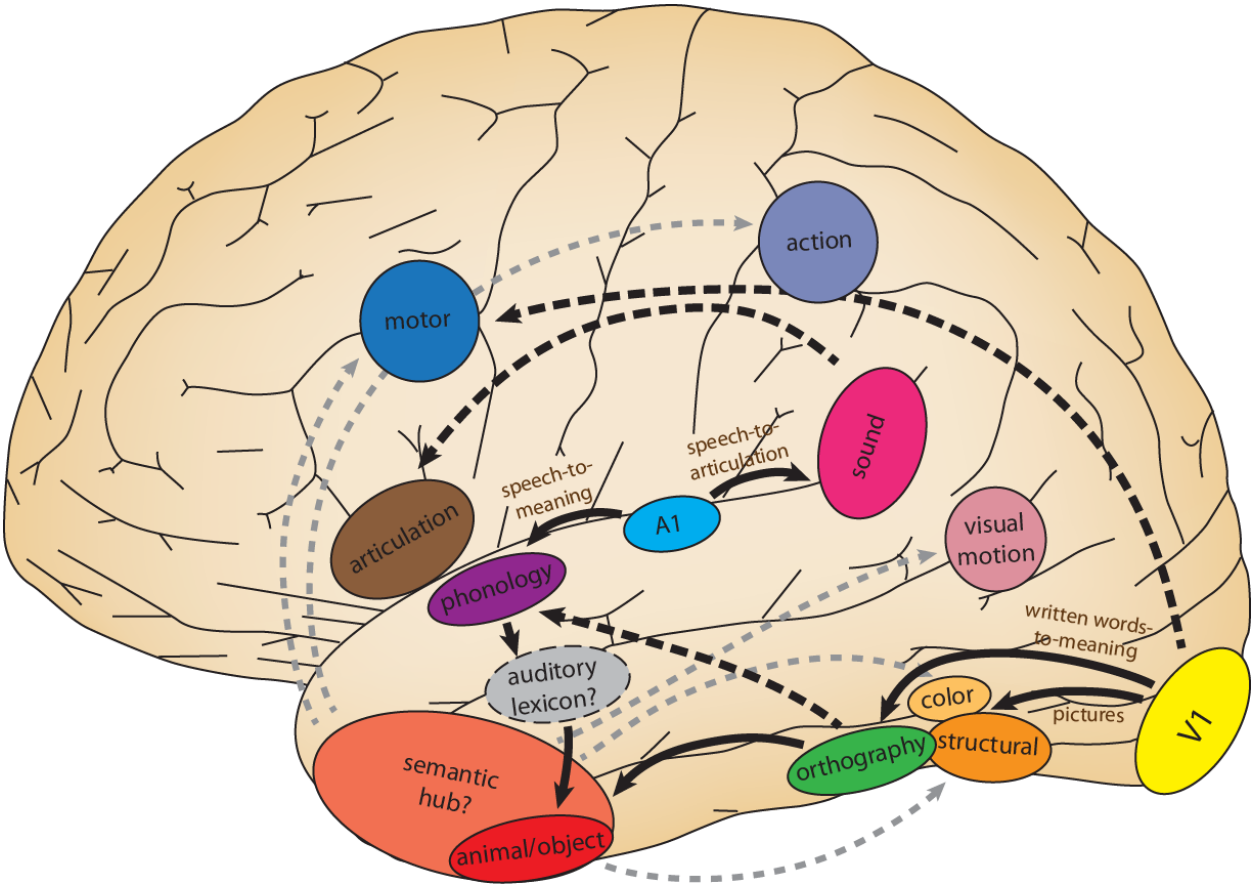

FIGURE V-2: FUNCTIONAL NEUROANATOMICAL MODEL OF WORD PROCESSING

This model of word processing suggests that auditory speech information is split into two pathways, the anterior/ventral pathway that maps sound onto meaning and moves through aSTG and STS while the dorsal stream maps sound onto articulation and location and moves posteriorly through Wernicke's area to inferior frontal lobe. Visual words enter through primary visual cortex and have orthographic information extracted by a potential visual word form area. It is also possible that some aspects of word processing occur within the dorsal visual stream. Visual orthographic elements may be phonologically recoded before lexical access can occur. Lexical access results in semantic access that involves anteroventral temporal lobe that may act as a semantic hub. This hub coordinates distributed domain-specific semantic information (e.g. information regarding structure, color, motor/functional attributes, auditory features, or visual motion) on a task-specific basis (gray lines).

The anterior STG is present within the auditory "what" stream and represents complex speech-specific sounds at the level of single units. This area has many of the same properties as the VWFA (e.g. prelexicality, invariance to speaker/font), and may in fact be the auditory analog for word processing (auditory word form area). The information from the

anterior/ventral stream continues to form an auditory word representation, yet whether this auditory lexicon is localized or distributed is unclear. The left STS also appears important for high-level speech feature representation, and may be the next processing stage for the information observed in aSTG (Warren et al., 2006, Liebenthal et al., 2005). The auditory representation of full words may potentially be represented by single units, although further evidence is necessary to strongly make this claim. While the dorsal pathway has traditionally been viewed as the main speech processing pathway of the human brain (Wernicke, 1874, Geschwind and Levitsky, 1968), the model presented here would suggest that the anteroventral pathway may be more important for the mapping of speech onto meaning.

Written word perception, on the other hand, necessarily starts in primary visual cortex. Visual processing typically is split into a ventral "what" stream and a dorsal "where" stream, although most previous studies have focused on the ventral visual areas for word processing. Orthographic information is extracted from low-level visual features, and these sublexical orthographic representations may be housed in the so-called visual word form area of the posterior fusiform gyrus (Dehaene et al., 2002). Whether the dorsal pathway is necessary for visual word processing is unclear, although several studies have suggested that it plays a role in recognizing words and strings of letters that are outside of normal formatting or presentation (Cohen et al., 2008, Vinckier et al., 2007). Given the role of the dorsal-speech pathway for mapping sound to articulation, one might hypothesize that the dorsal visual stream is important for mapping visual words to writing representations. In fact, one study demonstrated visual presentation of letters activated premotor areas involved in writing (Longcamp et al., 2003). Regardless of how much the dorsal visual stream contributes to word recognition, most current evidence suggests that the VWFA is the main brain region involved in

the processing of written words, and this fits with the idea that the ventral stream maps visual inputs to meaning.

How visual orthographic information eventually interfaces with the mental lexicon from this point forward is unknown. The evidence presented in Chapter IV suggests that phonological access occurs during visual word processing, even for tasks that do not explicitly require phonological proceeding, but it is not clear whether this is necessary for lexical access or not. Therefore, othorgraphic information may directly yield lexical recognition, or may first proceed through phonological routes to access word identity.

Regardless of the details of phonological recording, both spoken and written word representations eventually interface with the underlying semantic representation. The anteroventral temporal lobe is clearly important for semantic knowledge retrieval and the supramodal category-specific activity seen with the depth and laminar electrodes may in fact be a manifestation of the categorical organization of the aTL semantic hub. Semantic knowledge retrieval, with or without a hub, relies on the activity of distributed brain regions that may each contribute different domain-specific knowledge to the full instantiation of the particular semantic concept. Furthermore, the precise representations that are activated may be highly task specific. For example, when accessing the knowledge regarding concrete objects, if the task demands structural/visual knowledge, the posterior occipital-temporal region in the ventral visual stream may be activated as suggested in Section III.4. Or, if knowledge regarding the functional associations of tool use is necessary, motor areas in the frontal lobe may be accessed. This hypothesis would explain the apparent task dependence of seeing posterior occipital-temporal activity differentiating animals and object when presented with written words opposed to pictures (Devlin et al., 2005).

Overall, this model is consistent with the results presented in this dissertation and much of the existing literature regarding word processing. While the work presented here has examined specific aspects of semantic representations, and auditory/visual word processing, further research is necessary to fully understand the flow of sub-lexical, lexical, and semantic information required for language processing.

2. APPLICATIONS OF LANGUAGE DECODING

In this dissertation, we have utilized multivariate machine learning techniques to demonstrate that we are able to extract a wide variety of language-related information from electrophysiological information. The ability of support vector machines to handle high-dimensional data allowed for the decoding of semantic category and individual words from dense arrays of EEG and MEG sensors, and examination of highly distributed representations (Chapter II). The use of these classification techniques also allowed for a quantification of the amounts of information in various modalities of neurophysiological recordings such as EEG vs. MEG, or MUA vs. CSD. Finally, we were able to use decoding algorithms to demonstrate the time-course of information present in these recordings, showing early semantic information in avTL, and the temporal extraction of phonetic information from single unit recordings. However, besides simply studying the neural basis for language, the decoding of language information has other potential practical applications.

2.1. FEASIBILITY AND PERFORMANCE

Over the last several years, interest in decoding brain activity has increased substantially (Kellis et al., 2010, Bradberry et al., 2010, Chadwick et al., 2010, Formisano et al., 2008, Haxby et al., 2001a, Kamitani and Tong, 2005, Kay et al., 2008, Kim et al., 2008, Liu et al.,

2009, Pasley et al., 2012, Donoghue et al., 2007, Hochberg et al., 2006, Brumberg et al., 2009, Guenther et al., 2009), and many studies of neural decoding have received increasing attention by the popular press. While in some cases, the decoding is sufficiently accurate for practical purposes (e.g. single unit motor imagery decoding), in many cases, decoding is only viable as a novel technique to explore underlying neural information in a multivariate fashion. The results presented here largely fall into the latter category, although with additional research and study, language decoding may approach practical levels of accuracy.

While in this dissertation, we mainly utilize machine learning techniques for the study of language representations, this work may provide a foundation by which language decoding for communication prostheses may be viable in the future. The work reported here suggests that while language information can be extracted from electrophysiological recordings, further work is needed to improve the performance of such a language prosthetic device. Ideally, adequate performance from the decoding of a single trial would be required for a viable language decoding device.

One method that may improve classifier performance would be to employ more aggressive feature selection. In Chapter II, the SVM and computed features were not tuned to maximize performance, but instead allowed for the utilization of information from *all* sensors and times to investigate the spatiotemporal representation of semantic information. While SVMs generate a sparse model for classification that helps prevent overfitting, very high dimensional feature spaces can often prevent sufficient sparsity when using a small training set. Therefore, by limiting the sensor selection to particular regions of interest (such as ventral temporal areas, inferior frontal areas, and posterior temporal occipital areas), the number of considered features can be reduced while maintaining a majority of the available information.

The other way to prevent overfitting would be to simply increase the size of the training set, a task that is often challenging due to constraints on subjects' time and patience.

While overfitting may reduce the generalization performance of decoding algorithms, it is perhaps more important to ensure that the computed features carry as much language information as possible. Optimizing features for any decoding task is not a trivial process, and entirely domain-specific. In this dissertation, we have explored time-domain features for the decoding of semantic information from extracranial EEG and MEG, frequency-domain information in the form of gamma-band activity and multi-unit activity for decoding semantic category in depth electrodes, and firing rate activity for the decoding of complex acoustic/phonetic information from single unit recordings. From the results of the preceding chapters, one might conclude that information based on unit-firing measures (e.g. gamma-band activity, MUA, and single unit firing) perform the best. We have yet to explore features based on measures of synchrony (e.g. correlation, coherence), phase, or spike timing for single unit recordings. Studies have suggested that coherence measures may be important for discriminating grammatical classes (e.g. nouns vs. verbs) (Khader and Rosler, 2004, Weiss and Mueller, 2003), that the phase of oscillations in the brain is important for auditory perception (Ding and Simon, 2012, Saoud et al., 2012), and a spike timing may be as informative as spike firing rate (Mehta et al., 2002, Huxter et al., 2003). It is important that these features are explored in the future with regard to contributing to decoding performance of language information.

For any communication device based on the decoding of language information, it is important that the available vocabulary be large enough for sufficient expressive power. In the English language, it has been shown that the 1,000 most frequent words provide coverage of

approximately 75% of a variety of text corpora, 6,000 words provide 90% coverage, and 15,000 words provide 95% coverage (Nation and Waring, 1997). These vocabulary sizes are likely overestimates of the number of words needed to communicate at the most basic levels. In fact, by looking at the text coverage for novels aimed at teenagers, it was found at approximately 3,000 words provides over 95% coverage (Hirsh and Nation, 1992). It is likely that even smaller vocabularies might be useful in particular contexts, however, the 10-word decoding in Chapter II is far from adequate even for limited tasks. While this dissertation relied heavily on the use of machine learning classifiers, these algorithms, by themselves, are not ideal for the decoding of large vocabularies. An algorithm that is highly extensible is necessary for the decoding of semantic information. In Section II.2.7, a hierarchical decoding method was presented in which object size was first decoded using one set of features, followed by the decoding of semantic category, and finally individual words. This cascaded structure would allow for the decoding of general categories using relevant information before extracting detailed word-level information. For example, one might imagine first the decoding of grammatical class (e.g. nouns, verbs), followed by the decoding of semantic category (e.g. animals, tools), followed by even finer distinctions (e.g. birds, mammals, insects). This could drastically narrow the semantic search space before attempting to resolve specific concepts. A number of other approaches may also allow for the use of larger vocabularies, but it is clear that this is a major hurdle for language decoding based on semantic properties.

2.2. DECODING MULTIPLE LEVELS OF LANGUAGE

Considering the level of language processing a potential communication prosthesis would attempt to decode is important. This choice is highly dependent on the patient population such a device would target. For a patient with an expressive aphasia, the

impairment exists at the level of generating coherent, fluent sentences, usually leaving the semantic system intact. Therefore, extracting information at the semantic/conceptual level may be required for this type of patient. On the other hand, for a patient with amyotrophic lateral sclerosis, the deficit comes at the level of the motor neurons. In such a case, the entire language system is intact except for the very last stage of execution of articulatory commands. In this case, while a semantic-level decoder would still work, the incorporation of low-level lexical, phonetic, and articulatory information could significantly improve the performance of a communication prosthesis.

Even with just the information examined in this dissertation, it may be possible to elegantly merge phonetic information with semantic information to obtain higher word decoding performance. Automatic speech recognition (ASR) algorithms do this well, and it is possible such a solution may work for the decoding of language information from the brain. In ASR paradigms, the phonetics are modeled with Hidden Markov Models (HMMs) from which language models, or word transition probabilities, influence the likelihood of a particular word given the past history of words (Church and Hanks, 1990). In the case of neural data, one could treat the firing rates from individual neurons (if microelectrodes are used) or gamma-band power (from macroelectrodes) that carry phonetic information as the observed output variables of an HMM. Using a separate set of semantically relevant features, one could estimate probabilities of various words based on their semantic class. These probabilities can be treated as a prior probability, and combining this with the estimated HMM probability would yield a posterior probability for individual words within the set vocabulary. This method of combining phonetic and semantic information might improve decoding performance and may provide another extensible framework by which larger vocabularies can be used.

As a proof of principle, this technique for decoding individual words was tested on phonetic information from the patient discussed in Chapter IV as well as semantic information from one of the patients with implanted laminar electrodes (L2) as seen in Chapter III. A HMM trained on high gamma power (HGP) from intracranial surface electrodes over posterior temporal-parietal junction as well as firing activity from unit 6a allowed for the decoding of phonological information from a time-course of HGP and unit firing activity (Figure V-3B). Therefore, given a set of input features Y, the HMM provides the likelihood of these features given each of ten possible repeated words from the SA task, $P(Y|\text{word})$, by which the word with the largest likelihood could be chosen as the decoded word. Utilizing this phonological information alone yielded an accuracy of 41%. A Naïve Bayes Classifier trained to decode animal/object category from MUA information in patient L5 generates probabilities that a given trial is a word referring to an animal or object (Figure V-3C). These probabilities can be used to compute the probabilities of each individual word by assuming a uniform distribution within each category.

$$P(\text{claw}) = P(\text{cricket}) = \cdots = P(\text{lion}) = \frac{P(\text{animal})}{5}$$

$$P(\text{medal}) = P(\text{shelf}) = \cdots = P(\text{fork}) = \frac{P(\text{object})}{5} = \frac{1 - P(\text{animal})}{5}$$

By incorporating these probabilities into the HMM model, we can, in a sense, compute the joint probabilities, $P(Y|\text{word})P(\text{word}) = P(\text{word}, Y)$ which improves our decoding accuracy. In this case, incorporating the semantic information from the MUA features of patient L2 yields a final accuracy of 69% (versus 41% for phonological information alone). While this proof-of-principle uses data from two patients, which is not a realistic scenario for accurately judging the performance of such a technique, it does show that combining

information in this way can lead to higher decoding accuracies. In the future, utilizing language information from various stages of processing could allow for high performing word decoding from neural signals.

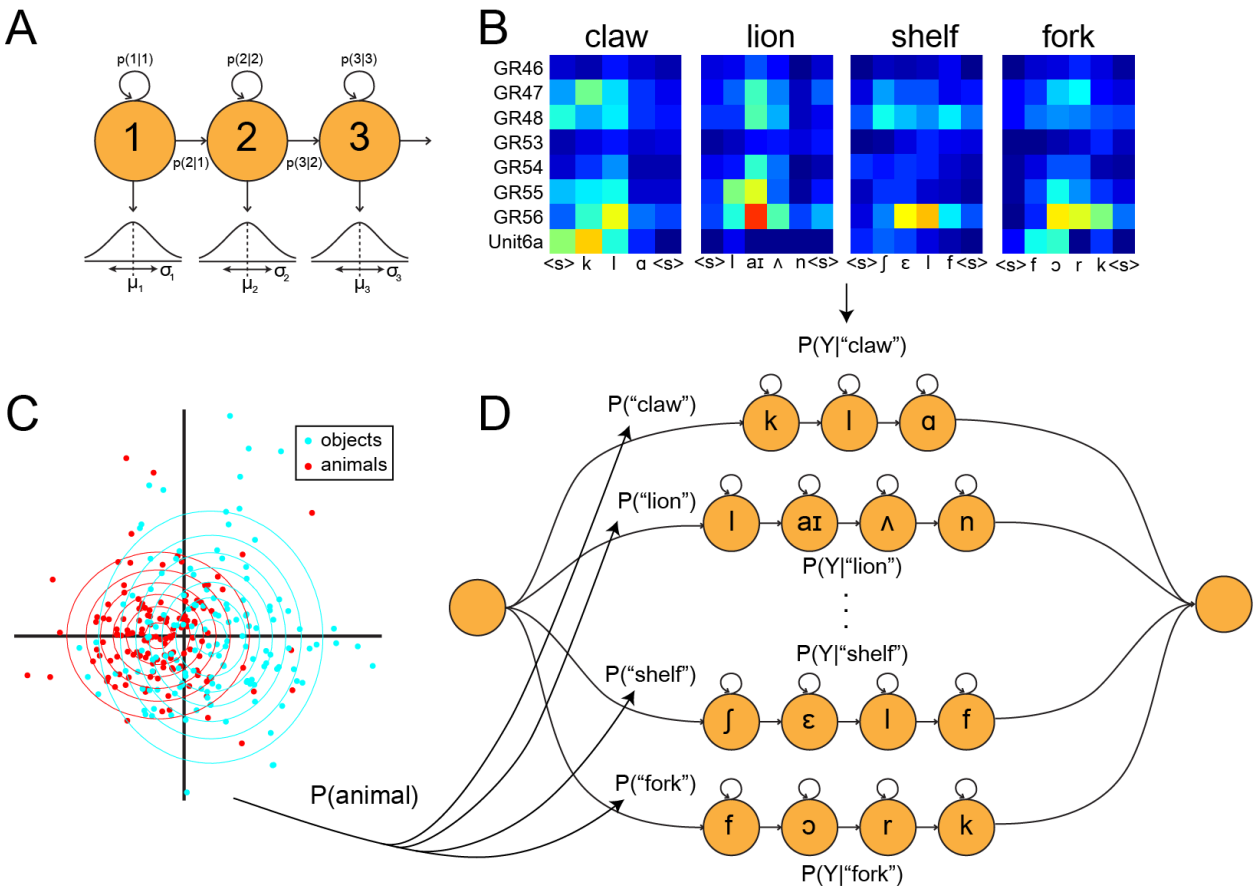

FIGURE V-3: COMBINATION OF PHONOLOGICAL AND SEMANTIC INFORMATION USING HIDDEN MARKOV MODELS

A) An example of a Hidden Markov Model that has hidden states 1, 2 and 3 that emit observed features that are characterized by a mean and standard deviation. Transitions between states are given by transition probabilities of moving from one state to the next, for remaining in the same state. This type of model is commonly used in automatic speech recognition, and can be used to model the phonological information in a time-course of high gamma power or unit firing. B) The mean high gamma power features (computed over a set of surface electrodes) as well as firing rate for unit 6a for the trained HMM model for each of four example words. Each phoneme would represent a single state that, on average, corresponds to the HGP and firing rate measures as shown. These features were computed for the patient described in Chapter IV. C) MUA features computed for patient L5 in Chapter III can be decoded by a Naïve Bayes Classifier (NBC). Trials are represented as points on a 2-dimensional plane. For each trial the NBC generates the probabilities $P(\text{animal})$ and $P(\text{object}) = 1 - P(\text{animal})$. These probabilities can be used to generate probabilities for each of the 10 words by assuming a uniform probability within each category. D)

(Figure V-3 continued) The HMM trained on the phonological features can use the word probabilities from the semantic NBC to improve decoding accuracy from 41% to 69%.

While this dissertation has utilized experiments that require language comprehension, a future language prosthetic device would aim to decode language production. In most models of language production, a single source of semantic knowledge underlies both comprehension and production (Martin, 2003, Patterson et al., 2007, Caramazza and Mahon, 2003). Therefore, it is reasonable to begin exploring the decoding of semantic knowledge from language comprehension experiments. It will eventually be necessary to understand the differences in the time course of activity when accessing semantic knowledge for production, yet it is likely that the spatial distribution of activity will be consistent. On the other hand, understanding the neural basis of articulation for particular phonemes would provide another source of relevant speech-specific information that could be extracted in patients with preserved motor planning ability. Several studies have examined the decoding of articulatory activity with relative success (Brumberg et al., 2009, Guenther, 2009, Guenther et al., 2009). Incorporating low-level motor planning activity with higher level lexico-semantic activity could prove to be a powerful set of features for decoding language intent.

One might also imagine that the same phonological lexicon is utilized for both speech comprehension and speech production. In Chapter IV, we demonstrated that speech-specific neurons fired in response to self-generated speech, although at a lower rate. This activity is most likely a response to the self-generated auditory input, however, a component of this activity may be due to phonological access for speech production. Studying the activity of such phonologically-tuned neurons during covert speech could potentially separate these two effects, and would provide an additional level of information for speech decoding. Several studies have indicated that the input and output phonological representations overlap, at least

partially, at the pSTG and STS (Buchsbaum et al., 2001, Price et al., 1996, Indefrey and Levelt, 2004, Liebenthal et al., 2005, Binder et al., 2000), consistent with the potential function of the dorsal auditory stream in which sound in mapped onto articulation. Investigation into whether the aSTG neurons described in Chapter IV, that are potentially in anterior/ventral auditory "what" stream, play an active role in speech production is an important next step.

2.3. RECORDING MODALITY

The recording modality of a potential language prosthetic device is important to consider. Several choices, such as MEG and fMRI, are not viable solutions simply due to their lack of portability. The main division between the available modalities is whether the electrodes are non-invasive or intracranially implanted. As seen in Chapter II, EEG electrodes provide excellent coverage of the cortical surface and can allow for the extraction of distributed semantic information. On the other hand, the spatial resolution and ability to more robustly record high frequency information in the intracranial macroelectrodes provided increased decoding performance given an appropriate recording location. Moving to penetrating microelectrodes provides the highest spatial resolution and the ability to record highly specific unit-firing activity. Therefore, the quality of the recorded information tended to improve as the electrodes became more invasive, at the cost of reduced cortical coverage.

It is unclear whether non-invasive recording techniques like EEG will be sufficient for a fully functional language-prosthetic device. However, if the severity of the communication deficit justifies the use of invasive recording techniques, a combination of microelectrodes for the recording of very specific phonetic/articulatory activity and macroelectrodes for recording more distributed lexico-semantic information may prove to work well. In the end, the choice of

modality will depend on a combination of the accuracy and effectiveness of the decoding given the available information and the severity of the communication impairment.

One final aspect of consideration and research when attempting to decode language information for practical purposes is the idea of patient-specific or patient-general classification. This is also highly affected by choice of modality due to the variable stability of the recordings. Single unit recordings are typically only stable for several hours due to shifting of microelectrodes with respect to cell bodies. This necessarily requires a retraining or adaptation of any generated classifier as the population of recorded units change. On the other hand, recording techniques that measure population activity are more stable over time. Chapter II demonstrated that a patient-general classifier could be trained on EEG data due to the large lead fields of extracranial electrodes. Despite this, adapting such a classifier with patient specific data improves classifier performance (Section II.2.6). Intracranial macroelectrodes tend to demonstrate fairly stable recordings, however these electrodes depend on precise placement and implantation if generalization across subjects is desired.

3. FUTURE DIRECTIONS

While this dissertation has explored a number of aspects of language processing using classification algorithms, many questions remain to be explored. Chapter II demonstrated that semantic representations are widely distributed, however the functional organization of semantic knowledge still requires examination. As discussed in Section I.1.5, whether semantic knowledge is organized categorically, as a weighting and summation of individual features, or a combination of both remains unclear (Caramazza et al., 1990, Caramazza and Mahon, 2003, Mummery et al., 1998, Patterson et al., 2007). To more thoroughly test the idea of a feature-based semantic space, one could first compute the distances between various semantic

concepts using n-gram co-occurrences, as done by Mitchell et al. (Mitchell et al., 2008), or use a variety of other semantic measures such as the Google Similarity Index (Budanitsky and Hirst, 2006, Cilibrasi and Vitanyi, 2007, Jiang and Conrath, 1997). Neural "distances" in response to these semantic concepts (presented as words or images) could subsequently be computed using distance measures between a computed set of neural features (e.g. gamma band power across a select set of electrodes) and these two sets of distance matrices could be compared. One would expect that a feature-based semantic organization would yield very similar distance measures, while a category-based organization might not show correlations between these two distance measures. By examining neural features computed from different regions of interest (e.g. ATL, lateral cortical areas, or posterior temporal-occipital areas), we would be able to examine the organization of different cortical areas.

It is also important to examine the fluidity of representations in the language processing system. While much of this dissertation has worked under the assumption that semantic, lexical, and phonetic representations are fixed on the timescales being examined, these representations may fluidly change over the course of hours or days, even at the level of single units (Quiroga et al., 2009). One potential method for the examination of potentially changing representations would be to train a decoding algorithm on early presentations of a set of words, and observe the performance of the algorithm over time. A similar technique would also be useful for the quantification of the stability of both neural representations and recording techniques for uses in language decoding. Furthermore, utilization of online learning algorithms would allow for the incremental update of classifier models and could also be useful both practically and for the study of plasticity of neural representations.

As discussed in section V.2.2, it is also important to understand which portions of the language comprehension pathway are also utilized for language production. To study phonological representations during speech production, it will be important to require covert, rather than overt, speech to prevent confounded auditory input. It is unclear whether a single phonological lexicon exists for both comprehension and production, and the recording of high level acoustic-phonetic single units during a production task may provide insights into this question. The use of covert speech is also very important because any patient in need of a communication prosthesis would be unable to overtly produce speech. Guenther et al. (2009) have demonstrated that it is possible to extract imagined phoneme production at the level of decoding articulation information from motor areas, and it would be important to demonstrate the feasibility of decoding higher level concepts from imagined language production.

The work in this dissertation has focused on single word comprehension, but it is important to study how the processing and production of entire phrases or sentences differs. Single words do not carry syntactic information and it is important to understand this for any model of language processing. The processing of syntax requires integrating information over longer time scales, and the continued use of electrophysiology would provide high resolution temporal information that may provide insights into parallel or feedback processes. Also, while Section IV.3.7 provided some evidence that the phonetic coding of spontaneous speech may utilize the same representations as those observed during single word processing, higher levels of language comprehension are expected to be significantly different. Even at the acoustic-phonetic level, single words spoken in "citation form" carry significantly different auditory information than continuous speech; many phonemes and sounds are significantly reduced or omitted completely when speaking rapidly (Jurafsky and Martin, 2009). For many of the epilepsy patients who are admitted for electrode implantation and monitoring, continuous and

spontaneous conversations and electrophysiological data are recorded over the course of one to two weeks. This provides an enormously large dataset by which it may be possible to examine natural unconstrained speech perception and production.

Finally, one important future direction for the decoding of language information from neurophysiological recordings would be to demonstrate that this can be performed in real time. The work performed in this dissertation has exclusively been performed offline, and it is necessary to demonstrate that the speed of feature extraction and classification is sufficient to support online decoding. Furthermore, if a language prosthetic device is to be viable, it may also be necessary to investigate the tradeoff between computational complexity and performance of machine learning algorithms so that these techniques may be efficiently implemented on an embedded processor.

While many potential paths exist to extend this work, I believe that this dissertation has demonstrated that multivariate decoding and machine learning techniques are powerful methods for exploring multiple scales of neural processing and the extraction of language information from neurophysiological recordings. I hope that this work provides a foundation upon which further research into the spatiotemporal representation of language information can build upon. More importantly, I have confidence that given sufficient thought and research, a rapid, intuitive brain-based language prosthesis can be developed to significantly benefit patients who are unable to adequately communicate with the outside world.

VI. REFERENCES

ALLISON, T., MCCARTHY, G., NOBRE, A., PUCE, A. & BELGER, A. 1994. Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cereb Cortex,* 4**,** 544-54.

ALLISON, T., PUCE, A., SPENCER, D. D. & MCCARTHY, G. 1999. Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb Cortex,* 9**,** 415-30.

ARGUIN, M. 1996. Shape Integration for Visual Object Recognition and Its Implication in Category-Specific Visual Agnosia. *Visual Cognition,* 3**,** 221 - 276.

ARNOTT, S. R., BINNS, M. A., GRADY, C. L. & ALAIN, C. 2004. Assessing the auditory dual-pathway model in humans. *Neuroimage,* 22**,** 401-8.

BAESS, P., HORVATH, J., JACOBSEN, T. & SCHROGER, E. 2011. Selective suppression of self-initiated sounds in an auditory stream: An ERP study. *Psychophysiology,* 48**,** 1276-83.

BAKER, J. 1975. The DRAGON system--An overview. *Acoustics, Speech and Signal Processing, IEEE Transactions on,* 23**,** 24- 29.

BALDO, J. V. & SHIMAMURA, A. P. 1998. Letter and category fluency in patients with frontal lobe lesions. *Neuropsychology,* 12**,** 259-67.

BAR, M., KASSAM, K. S., GHUMAN, A. S., BOSHYAN, J., SCHMID, A. M., DALE, A. M., HAMALAINEN, M. S., MARINKOVIC, K., SCHACTER, D. L., ROSEN, B. R. & HALGREN, E. 2006. Top-down facilitation of visual recognition. *Proc Natl Acad Sci U S A,* 103**,** 449-54.

BARNEA, A. & BREZNITZ, Z. 1998. Phonological and orthographic processing of Hebrew words: electrophysiological aspects. *The Journal of genetic psychology,* 159**,** 492-504.

BARTHO, P., HIRASE, H., MONCONDUIT, L., ZUGARO, M., HARRIS, K. D. & BUZSAKI, G. 2004. Characterization of neocortical principal cells and interneurons by network interactions and extracellular features. *J Neurophysiol,* 92**,** 600-8.

BENTIN, S., KUTAS, M. & HILLYARD, S. A. 1993. Electrophysiological evidence for task effects on semantic priming in auditory word processing. *Psychophysiology,* 30**,** 161-9.

BENTIN, S., MOUCHETANT-ROSTAING, Y., GIARD, M. H., ECHALLIER, J. F. & PERNIER, J. 1999. ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. *J Cogn Neurosci,* 11**,** 235-60.

BINDER, J. R., FROST, J. A., HAMMEKE, T. A., BELLGOWAN, P. S., SPRINGER, J. A., KAUFMAN, J. N. & POSSING, E. T. 2000. Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex,* 10**,** 512-28.

BINDER, J. R., FROST, J. A., HAMMEKE, T. A., RAO, S. M. & COX, R. W. 1996. Function of the left planum temporale in auditory and linguistic processing. *Brain : a journal of neurology,* 119 (Pt 4)**,** 1239-47.

BINDER, J. R., LIEBENTHAL, E., POSSING, E. T., MEDLER, D. A. & WARD, B. D. 2004. Neural correlates of sensory and decision processes in auditory object identification. *Nat Neurosci,* 7**,** 295-301.

BINDER, J. R., MEDLER, D. A., WESTBURY, C. F., LIEBENTHAL, E. & BUCHANAN, L. 2006. Tuning of the human left fusiform gyrus to sublexical orthographic structure. *Neuroimage,* 33**,** 739-48.

BINDER, J. R. & MOHR, J. P. 1992. The topography of callosal reading pathways. A case-control analysis. *Brain : a journal of neurology,* 115 (Pt 6)**,** 1807-26.

BINNEY, R. J., EMBLETON, K. V., JEFFERIES, E., PARKER, G. J. & RALPH, M. A. 2010. The ventral and inferolateral aspects of the anterior temporal lobe are crucial in semantic memory: evidence from a novel direct comparison of distortion-corrected fMRI, rTMS, and semantic dementia. *Cereb Cortex,* 20**,** 2728-38.

BIRBAUMER, N. 2006. Breaking the silence: brain-computer interfaces (BCI) for communication and motor control. *Psychophysiology,* 43**,** 517-32.

BITTERMAN, Y., MUKAMEL, R., MALACH, R., FRIED, I. & NELKEN, I. 2008. Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. *Nature,* 451**,** 197-201.

BOATMAN, D., LESSER, R. P. & GORDON, B. 1995. Auditory speech processing in the left temporal lobe: an electrical interference study. *Brain and language,* 51**,** 269-90.

BOZEAT, S., LAMBON RALPH, M. A., PATTERSON, K., GARRARD, P. & HODGES, J. R. 2000. Non-verbal semantic impairment in semantic dementia. *Neuropsychologia,* 38**,** 1207-15.

BRADBERRY, T. J., GENTILI, R. J. & CONTRERAS-VIDAL, J. L. 2010. Reconstructing three-dimensional hand movements from noninvasive electroencephalographic signals. *J Neurosci,* 30**,** 3432-7.

BRAUN, M., HUTZLER, F., ZIEGLER, J. C., DAMBACHER, M. & JACOBS, A. M. 2009. Pseudohomophone effects provide evidence of early lexico-phonological processing in visual word recognition. *Hum Brain Mapp,* 30**,** 1977-89.

BROCA, P. 1861. Nouvelle observation d'aphe´mie produite par une le´sion de la troisie`me circonvolution frontale. *Bulletins de la Société d'anatomie (París),* 6**,** 398–407.

BRUMBERG, J. S., KENNEDY, P. R. & GUENTHER, F. H. Year. Artificial speech synthesizer control by brain-computer interface. *In:* Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009), September 6-10 2009 Brighton, U. K.: International Speech Communication Association.

BUCHSBAUM, B. R., HICKOK, G. & HUMPHRIES, C. 2001. Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science,* 25**,** 663-678.

BUDANITSKY, A. & HIRST, G. 2006. Evaluating WordNet-based Measures of Lexical Semantic Relatedness. *Computational Linguistics,* 32**,** 13-47.

BURWELL, R. D. 2000. The parahippocampal region: corticocortical connectivity. *Ann N Y Acad Sci,* 911**,** 25-42.

CALLAN, D. E., JONES, J. A., CALLAN, A. M. & AKAHANE-YAMADA, R. 2004. Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *Neuroimage,* 22**,** 1182-94.

CAMPBELL, P. K., JONES, K. E., HUBER, R. J., HORCH, K. W. & NORMANN, R. A. 1991. A silicon-based, three-dimensional neural interface: manufacturing processes for an intracortical electrode array. *Biomedical Engineering, IEEE Transactions on,* 38**,** 758-768.

CAPLAN, D. 2006. Why is Broca's area involved in syntax? *Cortex,* 42**,** 469-71.

CARAMAZZA, A. 1988. Some aspects of language processing revealed through the analysis of acquired aphasia: the lexical system. *Annu Rev Neurosci,* 11**,** 395-421.

CARAMAZZA, A., HILLIS, A. E., RAPP, B. C. & ROMANI, C. 1990. The multiple semantics hypothesis: Multiple confusions? *Cognitive Neuropsychology,* 7**,** 161 - 189.

CARAMAZZA, A. & MAHON, B. Z. 2003. The organization of conceptual knowledge: the evidence from category-specific semantic deficits. *Trends Cogn Sci,* 7**,** 354-361.

CARAMAZZA, A. & SHELTON, J. R. 1998. Domain-specific knowledge systems in the brain the animate-inanimate distinction. *J Cogn Neurosci,* 10**,** 1-34.

CASH, S. S., HALGREN, E., DEHGHANI, N., ROSSETTI, A. O., THESEN, T., WANG, C., DEVINSKY, O., KUZNIECKY, R., DOYLE, W., MADSEN, J. R., BROMFIELD, E., EROSS, L., HALASZ, P., KARMOS, G., CSERCSA, R., WITTNER, L. & ULBERT, I. 2009. The human K-complex represents an isolated cortical down-state. *Science,* 324**,** 1084-7.

CHADWICK, M. J., HASSABIS, D., WEISKOPF, N. & MAGUIRE, E. A. 2010. Decoding Individual Episodic Memory Traces in the Human Hippocampus. *Curr Biol*.

CHAN, A. M., BAKER, J. M., ESKANDAR, E., SCHOMER, D., ULBERT, I., MARINKOVIC, K., CASH, S. S. & HALGREN, E. 2011a. First-Pass Selectivity for Semantic Categories in Human Anteroventral Temporal Lobe. *J Neurosci,* 31**,** 18119-29.

CHAN, A. M., HALGREN, E., CARLSON, C., DEVINSKY, O., DOYLE, W., KUZNIECKY, R., THESEN, T., WANG, C., SCHOMER, D. L., ESKANDAR, E. & CASH, S. S. Year. Decoding multiscale word and category-specific spatiotemporal representations from intracranial EEG. *In:* Proceedings of the Computational and System Neuroscience Conference, 2010 Salt Lake City, UT.

CHAN, A. M., HALGREN, E., MARINKOVIC, K. & CASH, S. S. 2011b. Decoding word and category-specific spatiotemporal representations from MEG and EEG. *Neuroimage,* 54**,** 3028-39.

CHAN, A. M., SUN, F. T., BOTO, E. H. & WINGEIER, B. M. 2008. Automated seizure onset detection for accurate onset time determination in intracranial EEG. *Clin Neurophysiol,* 119**,** 2687-96.

CHANG, E. F., RIEGER, J. W., JOHNSON, K., BERGER, M. S., BARBARO, N. M. & KNIGHT, R. T. 2010. Categorical speech representation in human superior temporal gyrus. *Nat Neurosci,* 13**,** 1428-32.

CHAO, L. L., HAXBY, J. V. & MARTIN, A. 1999. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat Neurosci,* 2**,** 913-9.

CHAO, L. L. & MARTIN, A. 2000. Representation of manipulable man-made objects in the dorsal stream. *Neuroimage,* 12**,** 478-84.

CHEN, W. & UGURBIL, K. 1999. High spatial resolution functional magnetic resonance imaging at very-high-magnetic field. *Top Magn Reson Imaging,* 10**,** 63-78.

CHRISTOFFELS, I. K., FORMISANO, E. & SCHILLER, N. O. 2007. Neural correlates of verbal feedback processing: an fMRI study employing overt speech. *Hum Brain Mapp,* 28**,** 868-79.

CHURCH, K. W. & HANKS, P. 1990. Word association norms, mutual information, and lexicography. *Comput. Linguist.,* 16**,** 22-29.

CILIBRASI, R. & VITANYI, P. M. B. 2007. The Google Similarity Distance. *IEEE Trans Knowl Data Eng,* 19**,** 370-383.

CLARKE, A., TAYLOR, K. I. & TYLER, L. K. 2011. The evolution of meaning: spatio-temporal dynamics of visual object recognition. *J Cogn Neurosci,* 23**,** 1887-99.

COHEN, D. & CUFFIN, B. N. 1983. Demonstration of useful differences between magnetoencephalogram and electroencephalogram. *Electroencephalogr Clin Neurophysiol,* 56**,** 38-51.

COHEN, D., CUFFIN, B. N., YUNOKUCHI, K., MANIEWSKI, R., PURCELL, C., COSGROVE, G. R., IVES, J., KENNEDY, J. G. & SCHOMER, D. L. 1990. MEG versus EEG localization test using implanted sources in the human brain. *Ann Neurol,* 28**,** 811-7.

COHEN, L. & DEHAENE, S. 2004. Specialization within the ventral stream: the case for the visual word form area. *Neuroimage,* 22**,** 466-76.

COHEN, L., DEHAENE, S., NACCACHE, L., LEHERICY, S., DEHAENE-LAMBERTZ, G., HENAFF, M. A. & MICHEL, F. 2000. The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain,* 123 (Pt 2)**,** 291-307.

COHEN, L., DEHAENE, S., VINCKIER, F., JOBERT, A. & MONTAVONT, A. 2008. Reading normal and degraded words: contribution of the dorsal and ventral visual pathways. *Neuroimage,* 40**,** 353-66.

COHEN, L., LEHERICY, S., CHOCHON, F., LEMER, C., RIVAUD, S. & DEHAENE, S. 2002. Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain,* 125**,** 1054-69.

COLTHEART, M., CURTIS, B., ATKINS, P. & HALLER, M. 1993. *Models of reading aloud : dual-route and parallel-distributed-processing approaches,* Washington, DC, American Psychological Association.

CRAMMER, K. & SINGER, Y. 2002. On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of Machine Learning Research,* 2**,** 265-292.

CREUTZFELDT, O., OJEMANN, G. & LETTICH, E. 1989a. Neuronal activity in the human lateral temporal lobe. I. Responses to speech. *Exp Brain Res,* 77**,** 451-75.

CREUTZFELDT, O., OJEMANN, G. & LETTICH, E. 1989b. Neuronal activity in the human lateral temporal lobe. II. Responses to the subjects own voice. *Exp Brain Res,* 77**,** 476-89.

CRINION, J. T., LAMBON-RALPH, M. A., WARBURTON, E. A., HOWARD, D. & WISE, R. J. 2003. Temporal lobe regions engaged during normal speech comprehension. *Brain,* 126**,** 1193-201.

CRONE, N. E., BOATMAN, D., GORDON, B. & HAO, L. 2001a. Induced electrocorticographic gamma activity during auditory perception. Brazier Award-winning article, 2001. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology,* 112**,** 565-82.

CRONE, N. E., HAO, L., HART, J., JR., BOATMAN, D., LESSER, R. P., IRIZARRY, R. & GORDON, B. 2001b. Electrocorticographic gamma activity during word production in spoken and sign language. *Neurology,* 57**,** 2045-53.

CSERCSA, R., DOMBOVARI, B., FABO, D., WITTNER, L., EROSS, L., ENTZ, L., SOLYOM, A., RASONYI, G., SZUCS, A., KELEMEN, A., JAKUS, R., JUHOS, V., GRAND, L., MAGONY, A., HALASZ, P., FREUND, T. F., MAGLOCZKY, Z., CASH, S. S., PAPP, L., KARMOS, G., HALGREN, E. & ULBERT, I. 2010. Laminar analysis of slow wave activity in humans. *Brain,* 133**,** 2814-29.

CUFFIN, B. N. & COHEN, D. 1979. Comparison of the magnetoencephalogram and electroencephalogram. *Electroencephalogr Clin Neurophysiol,* 47**,** 132-46.

CURIO, G., NEULOH, G., NUMMINEN, J., JOUSMAKI, V. & HARI, R. 2000. Speaking modifies voice-evoked activity in the human auditory cortex. *Hum Brain Mapp,* 9**,** 183-91.

DALE, A. M., LIU, A. K., FISCHL, B. R., BUCKNER, R. L., BELLIVEAU, J. W., LEWINE, J. D. & HALGREN, E. 2000. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron,* 26**,** 55-67.

DAMASIO, H. & DAMASIO, A. R. 1980. The anatomical basis of conduction aphasia. *Brain : a journal of neurology,* 103**,** 337-50.

DAMASIO, H., GRABOWSKI, T. J., TRANEL, D., HICHWA, R. D. & DAMASIO, A. R. 1996. A neural basis for lexical retrieval. *Nature,* 380**,** 499-505.

DAMASIO, H., TRANEL, D., GRABOWSKI, T., ADOLPHS, R. & DAMASIO, A. 2004. Neural systems behind word and concept retrieval. *Cognition,* 92**,** 179-229.

DAVIES, R. R., GRAHAM, K. S., XUEREB, J. H., WILLIAMS, G. B. & HODGES, J. R. 2004. The human perirhinal cortex and semantic memory. *Eur J Neurosci,* 20**,** 2441-6.

DAVIS, S. & MERMELSTEIN, P. 1980. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *Acoustics, Speech and Signal Processing, IEEE Transactions on,* 28**,** 357-366.

DEHAENE, S. & COHEN, L. 2007. Cultural recycling of cortical maps. *Neuron,* 56**,** 384-98.

DEHAENE, S., COHEN, L., SIGMAN, M. & VINCKIER, F. 2005. The neural code for written words: a proposal. *Trends Cogn Sci,* 9**,** 335-41.

DEHAENE, S., LE CLEC, H. G., POLINE, J. B., LE BIHAN, D. & COHEN, L. 2002. The visual word form area: a prelexical representation of visual words in the fusiform gyrus. *Neuroreport,* 13**,** 321-5.

DEHAENE, S., NACCACHE, L., COHEN, L., BIHAN, D. L., MANGIN, J. F., POLINE, J. B. & RIVIERE, D. 2001. Cerebral mechanisms of word masking and unconscious repetition priming. *Nat Neurosci,* 4**,** 752-8.

DEHGHANI, N., CASH, S. S., ROSSETTI, A. O., CHEN, C. C. & HALGREN, E. 2010. Magnetoencephalography demonstrates multiple asynchronous generators during human sleep spindles. *J Neurophysiol,* 104**,** 179-88.

DEJERINE, J. 1892. Contribution a l'étude anatomo-pathologique et clinique des différentes variétés de cécité verbale. *Mémoires de la Société de Biologie,* 4**,** 61-90.

DELL, G. S. & O'SEAGHDHA, P. G. 1992. Stages of lexical access in language production. *Cognition,* 42**,** 287-314.

DELORME, A. & MAKEIG, S. 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods,* 134**,** 9-21.

DESAI, R., LIEBENTHAL, E., WALDRON, E. & BINDER, J. R. 2008. Left posterior temporal regions are sensitive to auditory categorization. *J Cogn Neurosci,* 20**,** 1174-88.

DESIMONE, R. 1991. Face-Selective Cells in the Temporal Cortex of Monkeys. *J Cogn Neurosci,* 3**,** 1-8.

DESIMONE, R., FLEMING, J. & GROSS, C. G. 1980. Prestriate afferents to inferior temporal cortex: an HRP study. *Brain Res,* 184**,** 41-55.

DEVLIN, J. T., MOORE, C. J., MUMMERY, C. J., GORNO-TEMPINI, M. L., PHILLIPS, J. A., NOPPENEY, U., FRACKOWIAK, R. S., FRISTON, K. J. & PRICE, C. J. 2002. Anatomic constraints on cognitive theories of category specificity. *Neuroimage,* 15**,** 675-85.

DEVLIN, J. T., RUSHWORTH, M. F. & MATTHEWS, P. M. 2005. Category-related activation for written words in the posterior fusiform is task specific. *Neuropsychologia,* 43**,** 69-74.

DHOND, R. P., WITZEL, T., DALE, A. M. & HALGREN, E. 2007. Spatiotemporal cortical dynamics underlying abstract and concrete word reading. *Hum Brain Mapp,* 28**,** 355-62.

DING, N. & SIMON, J. Z. 2012. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J Neurophysiol,* 107**,** 78-89.

DODD, B. & HERMELIN, B. 1977. Phonological coding by the prelinguistically deaf. *Attention, Perception, & Psychophysics,* 21**,** 413-417.

DONCHIN, E., SPENCER, K. M. & WIJESINGHE, R. 2000. The mental prosthesis: assessing the speed of a P300-based brain-computer interface. *IEEE Trans Rehabil Eng,* 8**,** 174-9.

DONOGHUE, J. P. 2008. Bridging the brain to the world: a perspective on neural interface systems. *Neuron,* 60**,** 511-21.

DONOGHUE, J. P., HOCHBERG, L. R., NURMIKKO, A. V., BLACK, M. J., SIMERAL, J. D. & FRIEHS, G. 2007. Neuromotor prosthesis development. *Med Health R I,* 90**,** 12-5.

DYKSTRA, A. R., CHAN, A. M., QUINN, B. T., ZEPEDA, R., KELLER, C. J., CORMIER, J., MADSEN, J. R., ESKANDAR, E. N. & CASH, S. S. 2012. Individualized localization and cortical surface-based registration of intracranial electrodes. *Neuroimage,* 59**,** 3563-3570.

EINEVOLL, G. T., PETTERSEN, K. H., DEVOR, A., ULBERT, I., HALGREN, E. & DALE, A. M. 2007. Laminar population analysis: estimating firing rates and evoked synaptic activity from multielectrode recordings in rat barrel cortex. *J Neurophysiol,* 97**,** 2174-90.

ELIADES, S. J. & WANG, X. 2005. Dynamics of auditory-vocal interaction in monkey auditory cortex. *Cereb Cortex,* 15**,** 1510-23.

ETCOFF, N. L., FREEMAN, R. & CAVE, K. R. 1991. Can we lose memories of faces? content specificity and awareness in a prosopagnosic. *J. Cognitive Neuroscience,* 3**,** 25-41.

FABO, D., MAGLOCZKY, Z., WITTNER, L., PEK, A., EROSS, L., CZIRJAK, S., VAJDA, J., SOLYOM, A., RASONYI, G., SZUCS, A., KELEMEN, A., JUHOS, V., GRAND, L., DOMBOVARI, B., HALASZ, P., FREUND, T. F., HALGREN, E., KARMOS, G. & ULBERT, I. 2008. Properties of in vivo interictal spike generation in the human subiculum. *Brain,* 131**,** 485-99.

FARWELL, L. A. & DONCHIN, E. 1988. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr Clin Neurophysiol,* 70**,** 510-23.

FIEBACH, C. J., FRIEDERICI, A. D., MULLER, K. & VON CRAMON, D. Y. 2002. fMRI evidence for dual routes to the mental lexicon in visual word recognition. *J Cogn Neurosci,* 14**,** 11-23.

FIEZ, J. A. & PETERSEN, S. E. 1998. Neuroimaging studies of word reading. *Proc Natl Acad Sci U S A,* 95**,** 914-21.

FLINKER, A., CHANG, E. F., KIRSCH, H. E., BARBARO, N. M., CRONE, N. E. & KNIGHT, R. T. 2010. Single-trial speech suppression of auditory cortex activity in humans. *J Neurosci,* 30**,** 16643-50.

FORMISANO, E., DE MARTINO, F., BONTE, M. & GOEBEL, R. 2008. "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science,* 322**,** 970-3.

FRANCIS, W. N. & KUCERA, H. 1982. *Frequency Analysis of English Usage: Lexicon and Grammar,* Boston, Houghton Mifflin.

FRIED, I., MACDONALD, K. A. & WILSON, C. L. 1997. Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron,* 18**,** 753-65.

FRISTON, K. J., HOLMES, A. P., WORSLEY, K. J., POLINE, J. P., FRITH, C. D. & FRACKOWIAK, R. S. J. 1994. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping,* 2**,** 189-210.

FROST, R. 1998. Toward a strong phonological theory of visual word recognition: true issues and false trails. *Psychol Bull,* 123**,** 71-99.

FUJITA, I., TANAKA, K., ITO, M. & CHENG, K. 1992. Columns for visual features of objects in monkey inferotemporal cortex. *Nature,* 360**,** 343-6.

GAINOTTI, G. 1996. Cognitive and Anatomical Locus of Lesion in a Patient with a Category-specific Semantic Impairment for Living Beings. *Cognitive Neuropsychology,* 13**,** 357-390.

GESCHWIND, N. & LEVITSKY, W. 1968. Human brain: left-right asymmetries in temporal speech region. *Science,* 161**,** 186-7.

GILBERT, A. L., REGIER, T., KAY, P. & IVRY, R. B. 2006. Whorf hypothesis is supported in the right visual field but not the left. *Proc Natl Acad Sci U S A,* 103**,** 489-94.

GONZALEZ ANDINO, S. L., GRAVE DE PERALTA, R., KHATEB, A., PEGNA, A. J., THUT, G. & LANDIS, T. 2007. A glimpse into your vision. *Human Brain Mapping,* 28**,** 614-624.

GRABOWSKI, T. J., DAMASIO, H. & DAMASIO, A. R. 1998. Premotor and prefrontal correlates of category-related lexical retrieval. *Neuroimage,* 7**,** 232-43.

GRAFTON, S. T., FADIGA, L., ARBIB, M. A. & RIZZOLATTI, G. 1997. Premotor cortex activation during observation and naming of familiar tools. *Neuroimage,* 6**,** 231-6.

GRILL-SPECTOR, K., KNOUF, N. & KANWISHER, N. 2004. The fusiform face area subserves face perception, not generic within-category identification. *Nat Neurosci,* 7**,** 555-62.

GRILL-SPECTOR, K., KOURTZI, Z. & KANWISHER, N. 2001. The lateral occipital complex and its role in object recognition. *Vision Res,* 41**,** 1409-22.

GRODZINSKY, Y. 1986. Language deficits and the theory of syntax. *Brain Lang,* 27**,** 135-59.

GRODZINSKY, Y., PINANGO, M. M., ZURIF, E. & DRAI, D. 1999. The critical role of group studies in neuropsychology: comprehension regularities in Broca's aphasia. *Brain Lang,* 67**,** 134-47.

GUENTHER, F. H. 2009. Real-time speech synthesis for neural prosthesis. *J Acoust Soc Am,* 125**,** 2496-2496.

GUENTHER, F. H., BRUMBERG, J. S., WRIGHT, E. J., NIETO-CASTANON, A., TOURVILLE, J. A., PANKO, M., LAW, R., SIEBERT, S. A., BARTELS, J. L., ANDREASEN, D. S., EHIRIM, P., MAO, H. & KENNEDY, P. R. 2009. A Wireless Brain-Machine Interface for Real-Time Speech Synthesis. *PLoS ONE,* 4**,** e8218.

HAFTING, T., FYHN, M., MOLDEN, S., MOSER, M. B. & MOSER, E. I. 2005. Microstructure of a spatial map in the entorhinal cortex. *Nature,* 436**,** 801-6.

HAGOORT, P., HALD, L., BASTIAANSEN, M. & PETERSSON, K. M. 2004. Integration of word meaning and world knowledge in language comprehension. *Science,* 304**,** 438-41.

HALGREN, E., BAUDENA, P., HEIT, G., CLARKE, J. M., MARINKOVIC, K. & CLARKE, M. 1994. Spatio-temporal stages in face and word processing. I. Depth-recorded potentials in the human occipital, temporal and parietal lobes [corrected]. *J Physiol Paris,* 88**,** 1-50.

HALGREN, E., DALE, A. M., SERENO, M. I., TOOTELL, R. B., MARINKOVIC, K. & ROSEN, B. R. 1999. Location of human face-selective cortex with respect to retinotopic areas. *Human Brain Mapping,* 7**,** 29-37.

HALGREN, E., DHOND, R. P., CHRISTENSEN, N., VAN PETTEN, C., MARINKOVIC, K., LEWINE, J. D. & DALE, A. M. 2002. N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *Neuroimage,* 17**,** 1101-16.

HALGREN, E., WANG, C., SCHOMER, D. L., KNAKE, S., MARINKOVIC, K., WU, J. & ULBERT, I. 2006. Processing stages underlying word recognition in the anteroventral temporal lobe. *Neuroimage,* 30**,** 1401-13.

HÄMÄLÄINEN, M. & ILMONIEMI, R. 1994. Interpreting magnetic fields of the brain: minimum norm estimates. *Medical and Biological Engineering and Computing,* 32**,** 35-42.

HANSON, V. L. & FOWLER, C. A. 1987. Phonological coding in word reading: evidence from hearing and deaf readers. *Memory & cognition,* 15**,** 199-207.

HART, J., JR. & GORDON, B. 1990. Delineation of single-word semantic comprehension deficits in aphasia, with anatomical correlation. *Ann Neurol,* 27**,** 226-31.

HAUK, O., DAVIS, M. H., KHERIF, F. & PULVERMULLER, F. 2008. Imagery or meaning? Evidence for a semantic origin of category-specific brain activity in metabolic imaging. *Eur J Neurosci,* 27**,** 1856-66.

HAXBY, J. V., GOBBINI, M. I., FUREY, M. L., ISHAI, A., SCHOUTEN, J. L. & PIETRINI, P. 2001a. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science,* 293**,** 2425-30.

HAXBY, J. V., GOBBINI, M. I., FUREY, M. L., ISHAI, A., SCHOUTEN, J. L. & PIETRINI, P. 2001b. Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science,* 293**,** 2425-2430.

HEINKS-MALDONADO, T. H., MATHALON, D. H., GRAY, M. & FORD, J. M. 2005. Fine-tuning of auditory cortex during speech production. *Psychophysiology,* 42**,** 180-90.

HEINKS-MALDONADO, T. H., NAGARAJAN, S. S. & HOUDE, J. F. 2006. Magnetoencephalographic evidence for a precise forward model in speech production. *Neuroreport,* 17**,** 1375-9.

HEIT, G., SMITH, M. E. & HALGREN, E. 1988. Neural encoding of individual words and faces by the human hippocampus and amygdala. *Nature,* 333**,** 773-5.

HICKOK, G. & POEPPEL, D. 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition,* 92**,** 67-99.

HICKOK, G. & POEPPEL, D. 2007. The cortical organization of speech processing. *Nat Rev Neurosci,* 8**,** 393-402.

HILLIS, A. E. & CARAMAZZA, A. 1991. Category-specific naming and comprehension impairment: a double dissociation. *Brain,* 114 (Pt 5)**,** 2081-94.

HIRANO, S., NAITO, Y., OKAZAWA, H., KOJIMA, H., HONJO, I., ISHIZU, K., YENOKURA, Y., NAGAHAMA, Y., FUKUYAMA, H. & KONISHI, J. 1997. Cortical activation by monaural speech sound stimulation demonstrated by positron emission tomography. *Exp Brain Res,* 113**,** 75-80.

HIRSH, D. & NATION, P. 1992. What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language,* 8**,** 689-96.

HOCHBERG, L. R., SERRUYA, M. D., FRIEHS, G. M., MUKAND, J. A., SALEH, M., CAPLAN, A. H., BRANNER, A., CHEN, D., PENN, R. D. & DONOGHUE, J. P. 2006. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature,* 442**,** 164-71.

HODGES, J. R., PATTERSON, K., OXBURY, S. & FUNNELL, E. 1992. Semantic dementia. Progressive fluent aphasia with temporal lobe atrophy. *Brain : a journal of neurology,* 115 (Pt 6)**,** 1783-806.

HOLCOMB, P. J. & NEVILLE, H. J. 1990. Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Lang Cog Proc,* 5**,** 281-312.

HOUDE, J. F., NAGARAJAN, S. S., SEKIHARA, K. & MERZENICH, M. M. 2002. Modulation of the auditory cortex during speech: an MEG study. *J Cogn Neurosci,* 14**,** 1125-38.

HOWARD, D., PATTERSON, K., WISE, R., BROWN, W. D., FRISTON, K., WEILLER, C. & FRACKOWIAK, R. 1992. The cortical localization of the lexicons. Positron emission tomography evidence. *Brain,* 115 (Pt 6)**,** 1769-82.

HOWARD, M. A., 3RD, VOLKOV, I. O., ABBAS, P. J., DAMASIO, H., OLLENDIECK, M. C. & GRANNER, M. A. 1996. A chronic microelectrode investigation of the tonotopic organization of human auditory cortex. *Brain Res,* 724**,** 260-4.

HUANG, M. X., SONG, T., HAGLER, D. J., JR., PODGORNY, I., JOUSMAKI, V., CUI, L., GAA, K., HARRINGTON, D. L., DALE, A. M., LEE, R. R., ELMAN, J. & HALGREN, E. 2007. A novel integrated MEG and EEG analysis method for dipolar sources. *Neuroimage,* 37**,** 731-48.

HUBEL, D. H. & WIESEL, T. N. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology,* 160**,** 106-54.

HUMPHREYS, G. W. & EVETT, L. J. 1985. Are there independent lexical and nonlexical routes in word processing? An evaluation of the dual-route theory of reading. *Behavioral and Brain Sciences,* 8**,** 689-705.

HUMPHREYS, G. W., EVETT, L. J. & TAYLOR, D. E. 1982. Automatic phonological priming in visual word recognition. *Mem Cognit,* 10**,** 576-90.

HUMPHREYS, G. W., RIDDOCH, M. J. & PRICE, C. J. 1997. Top-down processes in object identification: evidence from experimental psychology, neuropsychology and functional anatomy. *Philos Trans R Soc Lond B Biol Sci,* 352**,** 1275-82.

HUXTER, J., BURGESS, N. & O'KEEFE, J. 2003. Independent rate and temporal coding in hippocampal pyramidal cells. *Nature,* 425**,** 828-32.

INDEFREY, P. & LEVELT, W. J. 2004. The spatial and temporal signatures of word production components. *Cognition,* 92**,** 101-44.

INSAUSTI, R., AMARAL, D. G. & COWAN, W. M. 1987. The entorhinal cortex of the monkey: II. Cortical afferents. *J Comp Neurol,* 264**,** 356-95.

ISHAI, A., UNGERLEIDER, L. G., MARTIN, A., SCHOUTEN, J. L. & HAXBY, J. V. 1999. Distributed representation of objects in the human ventral visual pathway. *Proc Natl Acad Sci U S A,* 96**,** 9379-84.

JANCKE, L., WUSTENBERG, T., SCHEICH, H. & HEINZE, H. J. 2002. Phonetic perception and the temporal cortex. *Neuroimage,* 15**,** 733-46.

JEFFERIES, E., PATTERSON, K., JONES, R. W. & LAMBON RALPH, M. A. 2009. Comprehension of concrete and abstract words in semantic dementia. *Neuropsychology,* 23**,** 492-9.

JIANG, J. J. & CONRATH, D. W. Year. Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy. *In:* International Conference Research on Computational Linguistics (ROCLING X), 1997. 9008.

JOACHIMS, T. 1999. Making large-Scale SVM Learning Practical. *In:* SCHOLKOPF, B., BURGES, C. J. C. & SMOLA, A. J. (eds.) *Advances in Kernel Methods - Support Vector Learning.* Cambridge: MIT Press.

JOHNSON, J. S. & NEWPORT, E. L. 1989. Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cogn Psychol,* 21**,** 60-99.

JURAFSKY, D. & MARTIN, J. H. 2009. *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*, Pearson Prentice Hall.

KAMITANI, Y. & TONG, F. 2005. Decoding the visual and subjective contents of the human brain. *Nat Neurosci,* 8**,** 679-85.

KANWISHER, N., DOWNING, P., EPSTEIN, R. & KOURTZI, Z. 2001. Functional Neuroimaging of Visual Recognition. *In:* CABEZA, R. & KINGSTON, A. (eds.) *Handbook of Functional NeuroImaging of Cognition.* Cambridge, MA: MIT Press.

KAPER, M., MEINICKE, P., GROSSEKATHOEFER, U., LINGNER, T. & RITTER, H. 2004. BCI Competition 2003--Data set IIb: support vector machines for the P300 speller paradigm. *IEEE Trans Biomed Eng,* 51**,** 1073-6.

KAY, K. N., NASELARIS, T., PRENGER, R. J. & GALLANT, J. L. 2008. Identifying natural images from human brain activity. *Nature,* 452**,** 352-5.

KELLER, C. J., TRUCCOLO, W., GALE, J. T., ESKANDAR, E., THESEN, T., CARLSON, C., DEVINSKY, O., KUZNIECKY, R., DOYLE, W. K., MADSEN, J. R., SCHOMER, D. L., MEHTA, A. D., BROWN, E. N.,

HOCHBERG, L. R., ULBERT, I., HALGREN, E. & CASH, S. S. 2010. Heterogeneous neuronal firing patterns during interictal epileptiform discharges in the human cortex. *Brain : a journal of neurology,* 133**,** 1668-81.

KELLIS, S., MILLER, K., THOMSON, K., BROWN, R., HOUSE, P. & GREGER, B. 2010. Decoding spoken words using local field potentials recorded from the cortical surface. *Journal of Neural Engineering,* 7**,** 056007.

KHADER, P. & ROSLER, F. 2004. EEG power and coherence analysis of visually presented nouns and verbs reveals left frontal processing differences. *Neurosci Lett,* 354**,** 111-4.

KIM, D. H., VIVENTI, J., AMSDEN, J. J., XIAO, J., VIGELAND, L., KIM, Y. S., BLANCO, J. A., PANILAITIS, B., FRECHETTE, E. S., CONTRERAS, D., KAPLAN, D. L., OMENETTO, F. G., HUANG, Y., HWANG, K. C., ZAKIN, M. R., LITT, B. & ROGERS, J. A. 2010. Dissolvable films of silk fibroin for ultrathin conformal bio-integrated electronics. *Nat Mater*.

KIM, S. P., SIMERAL, J. D., HOCHBERG, L. R., DONOGHUE, J. P. & BLACK, M. J. 2008. Neural control of computer cursor velocity by decoding motor cortical spiking activity in humans with tetraplegia. *J Neural Eng,* 5**,** 455-76.

KIPKE, D. R., VETTER, R. J., WILLIAMS, J. C. & HETKE, J. F. 2003. Silicon-substrate intracortical microelectrode arrays for long-term recording of neuronal spike activity in cerebral cortex. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on,* 11**,** 151-155.

KREIMAN, G., KOCH, C. & FRIED, I. 2000. Category-specific visual responses of single neurons in the human medial temporal lobe. *Nat Neurosci,* 3**,** 946-53.

KUMAR, S., STEPHAN, K. E., WARREN, J. D., FRISTON, K. J. & GRIFFITHS, T. D. 2007. Hierarchical processing of auditory objects in humans. *PLoS Comput Biol,* 3**,** e100.

KUTAS, M. & FEDERMEIER, K. D. 2000. Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn Sci,* 4**,** 463-470.

KUTAS, M. & FEDERMEIER, K. D. 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu Rev Psychol,* 62**,** 621-47.

KUTAS, M. & HILLYARD, S. A. 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science,* 207**,** 203-5.

KUTAS, M. & HILLYARD, S. A. 1984. Brain potentials during reading reflect word expectancy and semantic association. *Nature,* 307**,** 161-3.

LACOURSE, J. R. & HLUDIK, F. C., JR. 1990. An eye movement communication-control system for the disabled. *IEEE Trans Biomed Eng,* 37**,** 1215-20.

LAMBON RALPH, M. A., LOWE, C. & ROGERS, T. T. 2007. Neural basis of category-specific semantic deficits for living things: evidence from semantic dementia, HSVE and a neural network model. *Brain,* 130**,** 1127-37.

LAMBON RALPH, M. A., POBRIC, G. & JEFFERIES, E. 2009. Conceptual knowledge is underpinned by the temporal pole bilaterally: convergent evidence from rTMS. *Cereb Cortex,* 19**,** 832-8.

LAMBON RALPH, M. A., SAGE, K., JONES, R. W. & MAYBERRY, E. J. 2010. Coherent concepts are computed in the anterior temporal lobes. *Proc Natl Acad Sci U S A,* 107**,** 2717-22.

LAPOINTE, L. L. 2005. *Aphasia and related neurogenic language disorders,* New York, Thieme.

LAVENEX, P. & AMARAL, D. G. 2000. Hippocampal-neocortical interaction: A hierarchy of associativity. *Hippocampus,* 10**,** 420-430.

LAVENEX, P., SUZUKI, W. A. & AMARAL, D. G. 2002. Perirhinal and parahippocampal cortices of the macaque monkey: projections to the neocortex. *J Comp Neurol,* 447**,** 394-420.

LENNEBERG, E. H. 1967. *Biological foundations of language,* Oxford.

LEUTHARDT, E. C., MILLER, K. J., SCHALK, G., RAO, R. P. & OJEMANN, J. G. 2006. Electrocorticography-based brain computer interface--the Seattle experience. *IEEE Trans Neural Syst Rehabil Eng,* 14**,** 194-8.

LEWIS, D. 1998. Naive (Bayes) at forty: The independence assumption in information retrieval Machine Learning: ECML-98. *In:* NÉDELLEC, C. & ROUVEIROL, C. (eds.). Springer Berlin / Heidelberg.

LIBERMAN, A. M. & MATTINGLY, I. G. 1989. A specialization for speech perception. *Science,* 243**,** 489-94.

LIEBENTHAL, E., BINDER, J. R., SPITZER, S. M., POSSING, E. T. & MEDLER, D. A. 2005. Neural substrates of phonemic perception. *Cereb Cortex,* 15**,** 1621-31.

LINDEN, H., PETTERSEN, K. H. & EINEVOLL, G. T. 2010. Intrinsic dendritic filtering gives low-pass power spectra of local field potentials. *J Comput Neurosci,* 29**,** 423-44.

LIU, A. K., DALE, A. M. & BELLIVEAU, J. W. 2002. Monte Carlo simulation studies of EEG and MEG localization accuracy. *Hum Brain Mapp,* 16**,** 47-62.

LIU, H., AGAM, Y., MADSEN, J. R. & KREIMAN, G. 2009. Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron,* 62**,** 281-90.

LONGCAMP, M., ANTON, J. L., ROTH, M. & VELAY, J. L. 2003. Visual presentation of single letters activates a premotor area involved in writing. *Neuroimage,* 19**,** 1492-500.

LOW, A., BENTIN, S., ROCKSTROH, B., SILBERMAN, Y., GOMOLLA, A., COHEN, R. & ELBERT, T. 2003. Semantic categorization in the human brain: spatiotemporal dynamics revealed by magnetoencephalography. *Psychol Sci,* 14**,** 367-72.

LUND, K. & BURGESS, C. 1996. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods,* 28**,** 203-208.

LUTZENBERGER, W., PULVERMULLER, F. & BIRBAUMER, N. 1994. Words and pseudowords elicit distinct patterns of 30-Hz EEG responses in humans. *Neurosci Lett,* 176**,** 115-8.

MAHON, B. Z. & CARAMAZZA, A. 2009. Concepts and categories: a cognitive neuropsychological perspective. *Annu Rev Psychol,* 60**,** 27-51.

MALMIVUO, J., SUIHKO, V. & ESKOLA, H. 1997. Sensitivity distributions of EEG and MEG measurements. *IEEE Trans Biomed Eng,* 44**,** 196-208.

MALMIVUO, J. A. & SUIHKO, V. E. 2004. Effect of skull resistivity on the spatial resolutions of EEG and MEG. *Ieee Transactions on Biomedical Engineering,* 51**,** 1276-1280.

MARINKOVIC, K. 2004. Spatiotemporal dynamics of word processing in the human cortex. *Neuroscientist,* 10**,** 142-52.

MARINKOVIC, K., DHOND, R. P., DALE, A. M., GLESSNER, M., CARR, V. & HALGREN, E. 2003. Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron,* 38**,** 487-97.

MARIS, E. & OOSTENVELD, R. 2007. Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods,* 164**,** 177-90.

MARTIN-ELKINS, C. L. & HOREL, J. A. 1992. Cortical afferents to behaviorally defined regions of the inferior temporal and parahippocampal gyri as demonstrated by WGA-HRP. *J Comp Neurol,* 321**,** 177-92.

MARTIN, A. 2007. The representation of object concepts in the brain. *Annu Rev Psychol,* 58**,** 25-45.

MARTIN, A. & CHAO, L. L. 2001. Semantic memory and the brain: structure and processes. *Curr Opin Neurobiol,* 11**,** 194-201.

MARTIN, A., WIGGS, C. L., UNGERLEIDER, L. G. & HAXBY, J. V. 1996. Neural correlates of category-specific knowledge. *Nature,* 379**,** 649-52.

MARTIN, R. C. 2003. Language processing: functional organization and neuroanatomical basis. *Annu Rev Psychol,* 54**,** 55-89.

MCCARTHY, R. A. 1995. *Semantic Knowledge and Semantic Representations,* New York, Psychology Press.

MECHELLI, A., PRICE, C. J., FRISTON, K. J. & ISHAI, A. 2004. Where bottom-up meets top-down: neuronal interactions during perception and imagery. *Cereb Cortex,* 14**,** 1256-65.

MECHELLI, A., PRICE, C. J., NOPPENEY, U. & FRISTON, K. J. 2003. A dynamic causal modeling study on category effects: bottom-up or top-down mediation? *J Cogn Neurosci,* 15**,** 925-34.

MECHELLI, A., SARTORI, G., ORLANDI, P. & PRICE, C. J. 2006. Semantic relevance explains category effects in medial fusiform gyri. *Neuroimage,* 30**,** 992-1002.

MEHTA, M. R., LEE, A. K. & WILSON, M. A. 2002. Role of experience and oscillations in transforming a rate code into a temporal code. *Nature,* 417**,** 741-6.

MESULAM, M. M. 1998. From sensation to cognition. *Brain,* 121 (Pt 6)**,** 1013-52.

MILNER, A. D. & GOODALE, M. A. 2006. *The visual brain in action (2nd ed.)*, New York, NY, US: Oxford University Press.

MION, M., PATTERSON, K., ACOSTA-CABRONERO, J., PENGAS, G., IZQUIERDO-GARCIA, D., HONG, Y. T., FRYER, T. D., WILLIAMS, G. B., HODGES, J. R. & NESTOR, P. J. 2010. What the left and right anterior fusiform gyri tell us about semantic memory. *Brain,* 133**,** 3256-68.

MISHKIN, M., UNGERLEIDER, L. G. & MACKO, K. A. 1983. Object vision and spatial vision: two cortical pathways. *Trends Neurosci,* 6**,** 414-417.

MITCHELL, T. M., SHINKAREVA, S. V., CARLSON, A., CHANG, K. M., MALAVE, V. L., MASON, R. A. & JUST, M. A. 2008. Predicting human brain activity associated with the meanings of nouns. *Science,* 320**,** 1191-5.

MOORE, C. J. & PRICE, C. J. 1999. A functional neuroimaging study of the variables that generate category-specific object processing differences. *Brain,* 122 (Pt 5)**,** 943-62.

MORMANN, F., KORNBLITH, S., QUIROGA, R. Q., KRASKOV, A., CERF, M., FRIED, I. & KOCH, C. 2008. Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. *The Journal of neuroscience : the official journal of the Society for Neuroscience,* 28**,** 8865-72.

MOURAO-MIRANDA, J., BOKDE, A. L., BORN, C., HAMPEL, H. & STETTER, M. 2005. Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data. *Neuroimage,* 28**,** 980-95.

MUMMERY, C. J., PATTERSON, K., HODGES, J. R. & PRICE, C. J. 1998. Functional neuroanatomy of the semantic system: divisible by what? *J Cogn Neurosci,* 10**,** 766-77.

MUMMERY, C. J., PATTERSON, K., HODGES, J. R. & WISE, R. J. 1996. Generating 'tiger' as an animal name or a word beginning with T: differences in brain activation. *Proc Biol Sci,* 263**,** 989-95.

MUMMERY, C. J., PATTERSON, K., PRICE, C. J., ASHBURNER, J., FRACKOWIAK, R. S. & HODGES, J. R. 2000. A voxel-based morphometry study of semantic dementia: relationship between temporal lobe atrophy and semantic memory. *Ann Neurol,* 47**,** 36-45.

MUMMERY, C. J., PATTERSON, K., WISE, R. J., VANDENBERGHE, R., PRICE, C. J. & HODGES, J. R. 1999. Disrupted temporal lobe connections in semantic dementia. *Brain,* 122 (Pt 1)**,** 61-73.

NÄÄTÄNEN, R. & PICTON, T. 1987. The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. *Psychophysiology,* 24**,** 375-425.

NATION, P. & WARING, R. 1997. Vocabulary size, text coverage and word lists. *In:* MCCARTHY, M. & SCHMITT, N. (eds.) *Vocabulary: Description, Acquisition and Pedagogy.* Cambridge: Cambridge University Press.

NEWMAN, R. L. & CONNOLLY, J. F. 2004. Determining the role of phonology in silent reading using event-related brain potentials. *Brain Res Cogn Brain Res,* 21**,** 94-105.

NIELSEN, J. M. 1946. *Agnosia, apraxia, aphasia: Their value in cerebral localization,* New York, Hoeber.

NIR, Y., FISCH, L., MUKAMEL, R., GELBARD-SAGIV, H., ARIELI, A., FRIED, I. & MALACH, R. 2007. Coupling between neuronal firing rate, gamma LFP, and BOLD fMRI is related to interneuronal correlations. *Curr Biol,* 17**,** 1275-85.

NIZNIKIEWICZ, M. & SQUIRES, N. K. 1996. Phonological processing and the role of strategy in silent reading: behavioral and electrophysiological evidence. *Brain and language,* 52**,** 342-64.

NOBRE, A. C., ALLISON, T. & MCCARTHY, G. 1994. Word recognition in the human inferior temporal lobe. *Nature,* 372**,** 260-3.

NOPPENEY, U., PATTERSON, K., TYLER, L. K., MOSS, H., STAMATAKIS, E. A., BRIGHT, P., MUMMERY, C. & PRICE, C. J. 2007. Temporal lobe lesions and semantic impairment: a comparison of herpes simplex virus encephalitis and semantic dementia. *Brain,* 130**,** 1138-47.

NOPPENEY, U., PRICE, C. J., PENNY, W. D. & FRISTON, K. J. 2006. Two distinct neural mechanisms for category-selective responses. *Cereb Cortex,* 16**,** 437-45.

NUMMINEN, J., SALMELIN, R. & HARI, R. 1999. Subject's own speech reduces reactivity of the human auditory cortex. *Neurosci Lett,* 265**,** 119-22.

O'KEEFE, J. & NADEL, L. 1987. *The Hippocampus as a Cognitive Map,* Oxford, Oxford University Press.

OBLESER, J., BOECKER, H., DRZEZGA, A., HASLINGER, B., HENNENLOTTER, A., ROETTINGER, M., EULITZ, C. & RAUSCHECKER, J. P. 2006. Vowel sound extraction in anterior superior temporal cortex. *Hum Brain Mapp,* 27**,** 562-71.

OBLESER, J., LEAVER, A. M., VANMETER, J. & RAUSCHECKER, J. P. 2010. Segregation of vowels and consonants in human auditory cortex: evidence for distributed hierarchical organization. *Frontiers in psychology,* 1**,** 232.

PARKER, G. J., LUZZI, S., ALEXANDER, D. C., WHEELER-KINGSHOTT, C. A., CICCARELLI, O. & LAMBON RALPH, M. A. 2005. Lateralization of ventral and dorsal auditory-language pathways in the human brain. *Neuroimage,* 24**,** 656-66.

PASLEY, B. N., DAVID, S. V., MESGARANI, N., FLINKER, A., SHAMMA, S. A., CRONE, N. E., KNIGHT, R. T. & CHANG, E. F. 2012. Reconstructing speech from human auditory cortex. *PLoS biology,* 10**,** e1001251.

PATTERSON, K., NESTOR, P. J. & ROGERS, T. T. 2007. Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat Rev Neurosci,* 8**,** 976-87.

PAUS, T., PERRY, D. W., ZATORRE, R. J., WORSLEY, K. J. & EVANS, A. C. 1996. Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. *Eur J Neurosci,* 8**,** 2236-46.

PERANI, D., CAPPA, S. F., BETTINARDI, V., BRESSI, S., GORNO-TEMPINI, M., MATARRESE, M. & FAZIO, F. 1995. Different neural systems for the recognition of animals and man-made tools. *Neuroreport,* 6**,** 1637-41.

PERANI, D., DEHAENE, S., GRASSI, F., COHEN, L., CAPPA, S. F., DUPOUX, E., FAZIO, F. & MEHLER, J. 1996. Brain processing of native and foreign languages. *Neuroreport,* 7**,** 2439-44.

PERANI, D., SCHNUR, T., TETTAMANTI, M., GORNO-TEMPINI, M., CAPPA, S. F. & FAZIO, F. 1999. Word and picture matching: a PET study of semantic category effects. *Neuropsychologia,* 37**,** 293-306.

PETTERSEN, K. H., DEVOR, A., ULBERT, I., DALE, A. M. & EINEVOLL, G. T. 2006. Current-source density estimation based on inversion of electrostatic forward solution: effects of finite extent of neuronal activity and conductivity discontinuities. *J Neurosci Methods,* 154**,** 116-33.

PEYRACHE, A., DEHGHANI, N., ESKANDAR, E. N., MADSEN, J. R., ANDERSON, W. S., DONOGHUE, J. A., HOCHBERG, L. R., HALGREN, E., CASH, S. S. & DESTEXHE, A. 2012. Spatiotemporal dynamics of neocortical excitation and inhibition during human sleep. *Proc Natl Acad Sci U S A,* 109**,** 1731-6.

PFURTSCHELLER, G. & COOPER, R. 1975. Frequency dependence of the transmission of the EEG from cortex to scalp. *Electroencephalogr Clin Neurophysiol,* 38**,** 93-6.

PFURTSCHELLER, G. & NEUPER, C. 2001. Motor imagery and direct brain-computer communication. *Proc IEEE,* 89**,** 1123-1134.

PFURTSCHELLER, G., NEUPER, C., SCHLOGL, A. & LUGGER, K. 1998. Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters. *IEEE Trans Rehabil Eng,* 6**,** 316-25.

PHILLIPS, J. A., NOPPENEY, U., HUMPHREYS, G. W. & PRICE, C. J. 2002. Can segregation within the semantic system account for category-specific deficits? *Brain,* 125**,** 2067-80.

POBRIC, G., JEFFERIES, E. & LAMBON RALPH, M. A. 2010a. Category-specific versus category-general semantic impairment induced by transcranial magnetic stimulation. *Curr Biol,* 20**,** 964-8.

POBRIC, G., JEFFERIES, E. & RALPH, M. A. 2007. Anterior temporal lobes mediate semantic representation: mimicking semantic dementia by using rTMS in normal participants. *Proc Natl Acad Sci U S A,* 104**,** 20137-41.

POBRIC, G., JEFFERIES, E. & RALPH, M. A. 2010b. Amodal semantic representations depend on both anterior temporal lobes: evidence from repetitive transcranial magnetic stimulation. *Neuropsychologia,* 48**,** 1336-42.

PRICE, C., THIERRY, G. & GRIFFITHS, T. 2005. Speech-specific auditory processing: where is it? *Trends Cogn Sci,* 9**,** 271-6.

PRICE, C. J. & DEVLIN, J. T. 2003. The myth of the visual word form area. *Neuroimage,* 19**,** 473-81.

PRICE, C. J., NOPPENEY, U., PHILLIPS, J. & DEVLIN, J. T. 2003a. How is the fusiform gyrus related to category-specificity? *Cognitive Neuropsychology,* 20**,** 561 - 574.

PRICE, C. J., WINTERBURN, D., GIRAUD, A. L., MOORE, C. J. & NOPPENEY, U. 2003b. Cortical localisation of the visual and auditory word form areas: a reconsideration of the evidence. *Brain and language,* 86**,** 272-86.

PRICE, C. J., WISE, R. J., WARBURTON, E. A., MOORE, C. J., HOWARD, D., PATTERSON, K., FRACKOWIAK, R. S. & FRISTON, K. J. 1996. Hearing and saying. The functional neuro-anatomy of auditory word processing. *Brain,* 119 (Pt 3)**,** 919-31.

PROVERBIO, A. M., VECCHI, L. & ZANI, A. 2004. From orthography to phonetics: ERP measures of grapheme-to-phoneme conversion mechanisms in reading. *J Cogn Neurosci,* 16**,** 301-17.

PULVERMULLER, F. 2005. Brain mechanisms linking language and action. *Nat Rev Neurosci,* 6**,** 576-82.

PULVERMULLER, F., EULITZ, C., PANTEV, C., MOHR, B., FEIGE, B., LUTZENBERGER, W., ELBERT, T. & BIRBAUMER, N. 1996a. High-frequency cortical responses reflect lexical processing: an MEG study. *Electroencephalogr Clin Neurophysiol,* 98**,** 76-85.

PULVERMULLER, F., PREISSL, H., LUTZENBERGER, W. & BIRBAUMER, N. 1996b. Brain rhythms of language: nouns versus verbs. *Eur J Neurosci,* 8**,** 937-41.

QUIROGA, R. Q., KRASKOV, A., KOCH, C. & FRIED, I. 2009. Explicit Encoding of Multimodal Percepts by Single Neurons in the Human Brain. *Current Biology,* 19**,** 1308-1313.

QUIROGA, R. Q., KREIMAN, G., KOCH, C. & FRIED, I. 2008. Sparse but not 'Grandmother-cell' coding in the medial temporal lobe. *Trends Cogn Sci,* 12**,** 87-91.

QUIROGA, R. Q., REDDY, L., KREIMAN, G., KOCH, C. & FRIED, I. 2005. Invariant visual representation by single neurons in the human brain. *Nature,* 435**,** 1102-7.

RASCH, M. J., GRETTON, A., MURAYAMA, Y., MAASS, W. & LOGOTHETIS, N. K. 2008. Inferring spike trains from local field potentials. *J Neurophysiol,* 99**,** 1461-76.

RAUSCHECKER, J. P. 1997. Processing of complex sounds in the auditory cortex of cat, monkey, and man. *Acta oto-laryngologica. Supplementum,* 532**,** 34-8.

RAUSCHECKER, J. P. 1998. Cortical processing of complex sounds. *Curr Opin Neurobiol,* 8**,** 516-21.

RAUSCHECKER, J. P. & SCOTT, S. K. 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci,* 12**,** 718-24.

RIDDOCH, M. J., HUMPHREYS, G. W., COLTHEART, M. & FUNNELL, E. 1988. Semantic systems or system? Neuropsychological evidence re-examined. *Cognitive Neuropsychology,* 5**,** 3-25.

RISH, I. 2001. An empirical study of the naive Bayes classifier. *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence,* 3**,** 41–46.

ROMANSKI, L. M., TIAN, B., FRITZ, J., MISHKIN, M., GOLDMAN-RAKIC, P. S. & RAUSCHECKER, J. P. 1999. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci,* 2**,** 1131-6.

ROUSCHE, P. J. & NORMANN, R. A. 1998. Chronic recording capability of the Utah Intracortical Electrode Array in cat sensory cortex. *J Neurosci Methods,* 82**,** 1-15.

SACCHETT, C. & HUMPHREYS, G. W. 1992. Calling a squirrel a squirrel but a canoe a wigwam: a category-specific deficit for artefactual objects and body parts. *Cognitive Neuropsychology,* 9**,** 73-86.

SAHIN, N. T., PINKER, S., CASH, S. S., SCHOMER, D. L. & HALGREN, E. 2009. Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science,* In Press.

SALEEM, K. S., SUZUKI, W., TANAKA, K. & HASHIKAWA, T. 2000. Connections between anterior inferotemporal cortex and superior temporal sulcus regions in the macaque monkey. *J Neurosci,* 20**,** 5083-101.

SALEEM, K. S. & TANAKA, K. 1996. Divergent projections from the anterior inferotemporal area TE to the perirhinal and entorhinal cortices in the macaque monkey. *J Neurosci,* 16**,** 4757-75.

SALEEM, K. S., TANAKA, K. & ROCKLAND, K. S. 1993. Specific and columnar projection from area TEO to TE in the macaque inferotemporal cortex. *Cereb Cortex,* 3**,** 454-64.

SANTHANAM, G., RYU, S. I., YU, B. M., AFSHAR, A. & SHENOY, K. V. 2006. A high-performance brain-computer interface. *Nature,* 442**,** 195-8.

SANTHANAM, G., SAHANI, M., RYU, S. & SHENOY, K. 2004. An extensible infrastructure for fully automated spike sorting during online experiments. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference,* 6**,** 4380-4.

SAOUD, H., JOSSE, G., BERTASI, E., TRUY, E., CHAIT, M. & GIRAUD, A.-L. 2012. Brain–Speech Alignment Enhances Auditory Cortical Responses and Speech Perception. *The Journal of Neuroscience,* 32**,** 275-281.

SAUR, D., KREHER, B. W., SCHNELL, S., KUMMERER, D., KELLMEYER, P., VRY, M. S., UMAROVA, R., MUSSO, M., GLAUCHE, V., ABEL, S., HUBER, W., RIJNTJES, M., HENNIG, J. & WEILLER, C. 2008. Ventral and dorsal pathways for language. *Proc Natl Acad Sci U S A,* 105**,** 18035-40.

SCOTT, S. K., BLANK, C. C., ROSEN, S. & WISE, R. J. 2000. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain,* 123 Pt 12**,** 2400-6.

SCOTT, S. K., ROSEN, S., LANG, H. & WISE, R. J. 2006. Neural correlates of intelligibility in speech investigated with noise vocoded speech--a positron emission tomography study. *J Acoust Soc Am,* 120**,** 1075-83.

SEIDENBERG, M. S. 1985. The time course of phonological code activation in two writing systems. *Cognition,* 19**,** 1-30.

SELTZER, B. & PANDYA, D. N. 1978. Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Res,* 149**,** 1-24.

SHANNON, R. V., ZENG, F. G., KAMATH, V., WYGONSKI, J. & EKELID, M. 1995. Speech recognition with primarily temporal cues. *Science,* 270**,** 303-4.

SHAO, S. Y., SHEN, K. Q., ONG, C. J., WILDER-SMITH, E. P. & LI, X. P. 2009. Automatic EEG artifact removal: a weighted support vector machine approach with error correction. *IEEE Trans Biomed Eng,* 56**,** 336-44.

SHERIDAN, J. & HUMPHREYS, G. W. 1993. A verbal-semantic category-specific recognition impairment. *Cognitive Neuropsychology,* 10**,** 143-184.

SHINKAREVA, S. V., MASON, R. A., MALAVE, V. L., WANG, W., MITCHELL, T. M. & JUST, M. A. 2008. Using FMRI brain activation to identify cognitive states associated with perception of tools and dwellings. *PLoS ONE,* 3**,** e1394.

SHOEB, A., EDWARDS, H., CONNOLLY, J., BOURGEOIS, B., TREVES, S. T. & GUTTAG, J. 2004. Patient-specific seizure onset detection. *Epilepsy Behav,* 5**,** 483-98.

SMITH, M. E., STAPLETON, J. M. & HALGREN, E. 1986. Human medial temporal lobe potentials evoked in memory and language tasks. *Electroencephalography and clinical neurophysiology,* 63**,** 145-59.

SPITSYNA, G., WARREN, J. E., SCOTT, S. K., TURKHEIMER, F. E. & WISE, R. J. 2006. Converging language streams in the human temporal lobe. *J Neurosci,* 26**,** 7328-36.

STEINSCHNEIDER, M., NOURSKI, K. V., KAWASAKI, H., OYA, H., BRUGGE, J. F. & HOWARD, M. A., 3RD 2011. Intracranial Study of Speech-Elicited Activity on the Human Posterolateral Superior Temporal Gyrus. *Cereb Cortex*.

SUPPES, P. & HAN, B. 2000. Brain-wave representation of words by superposition of a few sine waves. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 8738-8743.

SUPPES, P., HAN, B., EPELBOIM, J. & LU, Z.-L. 1999. Invariance between subjects of brain wave representations of language. *Proceedings of the National Academy of Sciences of the United States of America,* 96**,** 12953-12958.

SUPPES, P., LU, Z.-L. & HAN, B. 1997. Brain wave recognition of words. *Proceedings of the National Academy of Sciences of the United States of America,* 94**,** 14965-14969.

SUZUKI, W., SALEEM, K. S. & TANAKA, K. 2000. Divergent backward projections from the anterior part of the inferotemporal cortex (area TE) in the macaque. *J Comp Neurol,* 422**,** 206-28.

SUZUKI, W. A. 1996. The anatomy, physiology and functions of the perirhinal cortex. *Curr Opin Neurobiol,* 6**,** 179-86.

SUZUKI, W. A. & AMARAL, D. G. 1994. Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. *J Comp Neurol,* 350**,** 497-533.

TALLERMAN, M. & GIBSON, K. R. (eds.) 2012. *The Oxford Handbook of Language Evolution,* Oxford.

TALLON-BAUDRY, C. & BERTRAND, O. 1999. Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn Sci,* 3**,** 151-162.

TANAKA, K. 1996. Inferotemporal cortex and object vision. *Annu Rev Neurosci,* 19**,** 109-39.

THEUNISSEN, F. E., DAVID, S. V., SINGH, N. C., HSU, A., VINJE, W. E. & GALLANT, J. L. 2001. Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network,* 12**,** 289-316.

THOMPSON-SCHILL, S. L., AGUIRRE, G. K., D'ESPOSITO, M. & FARAH, M. J. 1999. A neural basis for category and modality specificity of semantic knowledge. *Neuropsychologia,* 37**,** 671-6.

TIAN, B. & RAUSCHECKER, J. P. 1994. Processing of frequency-modulated sounds in the cat's anterior auditory field. *J Neurophysiol,* 71**,** 1959-75.

TIAN, B., RESER, D., DURHAM, A., KUSTOV, A. & RAUSCHECKER, J. P. 2001. Functional specialization in rhesus monkey auditory cortex. *Science,* 292**,** 290-3.

TOURVILLE, J. A., REILLY, K. J. & GUENTHER, F. H. 2008. Neural mechanisms underlying auditory feedback control of speech. *Neuroimage,* 39**,** 1429-43.

TRANEL, D., DAMASIO, H. & DAMASIO, A. R. 1997. A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia,* 35**,** 1319-27.

TRUCCOLO, W., DONOGHUE, J. A., HOCHBERG, L. R., ESKANDAR, E. N., MADSEN, J. R., ANDERSON, W. S., BROWN, E. N., HALGREN, E. & CASH, S. S. 2011. Single-neuron dynamics in human focal epilepsy. *Nat Neurosci,* 14**,** 635-41.

TURKELTAUB, P. E. & COSLETT, H. B. 2010. Localization of sublexical speech perception components. *Brain and language,* 114**,** 1-15.

TYLER, L. K. & MOSS, H. E. 2001. Towards a distributed account of conceptual knowledge. *Trends Cogn Sci,* 5**,** 244-252.

TYLER, L. K., MOSS, H. E., DURRANT-PEATFIELD, M. R. & LEVY, J. P. 2000. Conceptual structure and the structure of concepts: a distributed account of category-specific deficits. *Brain Lang,* 75**,** 195-231.

ULBERT, I., HALGREN, E., HEIT, G. & KARMOS, G. 2001. Multiple microelectrode-recording system for human intracortical applications. *J Neurosci Methods,* 106**,** 69-79.

ULBERT, I., HEIT, G., MADSEN, J., KARMOS, G. & HALGREN, E. 2004a. Laminar analysis of human neocortical interictal spike generation and propagation: current source density and multiunit analysis in vivo. *Epilepsia,* 45 Suppl 4**,** 48-56.

ULBERT, I., MAGLOCZKY, Z., EROSS, L., CZIRJAK, S., VAJDA, J., BOGNAR, L., TOTH, S., SZABO, Z., HALASZ, P., FABO, D., HALGREN, E., FREUND, T. F. & KARMOS, G. 2004b. In vivo laminar electrophysiology co-registered with histology in the hippocampus of patients with temporal lobe epilepsy. *Exp Neurol,* 187**,** 310-8.

UNGERLEIDER, L. G. & HAXBY, J. V. 1994. 'What' and 'where' in the human brain. *Curr Opin Neurobiol,* 4**,** 157-65.

VAN ESSEN, D. C. & MAUNSELL, J. H. R. 1983. Hierarchical organization and functional streams in the visual cortex. *Trends Neurosci,* 6**,** 370-375.

VAN HOESEN, G. W. 1982. The parahippocampal gyrus: New observations regarding its cortical connections in the monkey. *Trends Neurosci,* 5**,** 345-350.

VAN HOESEN, G. W. & PANDYA, D. N. 1975. Some connections of the entorhinal (area 28) and perirhinal (area 35) cortices of the rhesus monkey. I. Temporal lobe afferents. *Brain Res Cogn Brain Res,* 95**,** 1-24.

VANRULLEN, R. & THORPE, S. J. 2001. The time course of visual processing: from early perception to decision-making. *J Cogn Neurosci,* 13**,** 454-61.

VAPNIK, V. N. 1995. *The nature of statistical learning theory,* New York, Springer-Verlag.

VINCKIER, F., DEHAENE, S., JOBERT, A., DUBUS, J. P., SIGMAN, M. & COHEN, L. 2007. Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. *Neuron,* 55**,** 143-56.

VINCKIER, F., NACCACHE, L., PAPEIX, C., FORGET, J., HAHN-BARMA, V., DEHAENE, S. & COHEN, L. 2006. "What" and "where" in word reading: ventral coding of written words revealed by parietal atrophy. *J Cogn Neurosci,* 18**,** 1998-2012.

VISSER, M., EMBLETON, K. V., JEFFERIES, E., PARKER, G. J. & RALPH, M. A. 2010a. The inferior, anterior temporal lobes and semantic memory clarified: novel evidence from distortion-corrected fMRI. *Neuropsychologia,* 48**,** 1689-96.

VISSER, M., JEFFERIES, E. & LAMBON RALPH, M. A. 2010b. Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. *J Cogn Neurosci,* 22**,** 1083-94.

VOULOUMANOS, A., KIEHL, K. A., WERKER, J. F. & LIDDLE, P. F. 2001. Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *J Cogn Neurosci,* 13**,** 994-1005.

WAI KEI, L., MING CHUI, D., JUN, S. & BIN BIN, F. Year. Automatic ECG interpretation via morphological feature extraction and SVM inference nets. *In:* Circuits and Systems, 2008. APCCAS 2008. IEEE Asia Pacific Conference on, Nov. 30 2008-Dec. 3 2008 2008. 254-258.

WARREN, J. D., SCOTT, S. K., PRICE, C. J. & GRIFFITHS, T. D. 2006. Human brain mechanisms for the early analysis of voices. *Neuroimage,* 31**,** 1389-97.

WARRINGTON, E. K. 1975. The selective impairment of semantic memory. *The Quarterly journal of experimental psychology,* 27**,** 635-57.

WARRINGTON, E. K. 1981. Concrete word dyslexia. *British journal of psychology,* 72**,** 175-96.

WARRINGTON, E. K. & MCCARTHY, R. 1983. Category specific access dysphasia. *Brain,* 106 (Pt 4)**,** 859-78.

WARRINGTON, E. K. & SHALLICE, T. 1980. Word-form dyslexia. *Brain : a journal of neurology,* 103**,** 99-112.

WARRINGTON, E. K. & SHALLICE, T. 1984. Category specific semantic impairments. *Brain,* 107 (Pt 3)**,** 829-54.

WAYDO, S., KRASKOV, A., QUIAN QUIROGA, R., FRIED, I. & KOCH, C. 2006. Sparse representation in the human medial temporal lobe. *The Journal of neuroscience : the official journal of the Society for Neuroscience,* 26**,** 10232-4.

WEISS, S. & MUELLER, H. M. 2003. The contribution of EEG coherence to the investigation of language. *Brain Lang,* 85**,** 325-43.

WERNICKE, C. 1874. *Der aphasische Symptomencomplex: eine psychologische Studie auf anatomischer Basis*, Cohn & Weigert.

WESSINGER, C. M., VANMETER, J., TIAN, B., VAN LARE, J., PEKAR, J. & RAUSCHECKER, J. P. 2001. Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *J Cogn Neurosci,* 13**,** 1-7.

WHALEN, D. H., BENSON, R. R., RICHARDSON, M., SWAINSON, B., CLARK, V. P., LAI, S., MENCL, W. E., FULBRIGHT, R. K., CONSTABLE, R. T. & LIBERMAN, A. M. 2006. Differentiation of speech and nonspeech processing within primary auditory cortex. *The Journal of the Acoustical Society of America,* 119**,** 575-81.

WHATMOUGH, C., CHERTKOW, H., MURTHA, S. & HANRATTY, K. 2002. Dissociable brain regions process object meaning and object structure during picture naming. *Neuropsychologia,* 40**,** 174-86.

WILSON, J. A., FELTON, E. A., GARELL, P. C., SCHALK, G. & WILLIAMS, J. C. 2006. ECoG factors underlying multimodal control of a brain-computer interface. *IEEE Trans Neural Syst Rehabil Eng,* 14**,** 246-50.

WINAWER, J., WITTHOFT, N., FRANK, M. C., WU, L., WADE, A. R. & BORODITSKY, L. 2007. Russian blues reveal effects of language on color discrimination. *Proc Natl Acad Sci U S A,* 104**,** 7780-5.

WOLTERS, C. H., ANWANDER, A., TRICOCHE, X., WEINSTEIN, D., KOCH, M. A. & MACLEOD, R. S. 2006. Influence of tissue conductivity anisotropy on EEG/MEG field and return current computation in a realistic head model: a simulation and visualization study using high-resolution finite element modeling. *Neuroimage,* 30**,** 813-26.

YVERT, G., PERRONE-BERTOLOTTI, M., BACIU, M. & DAVID, O. 2012. Dynamic Causal Modeling of Spatiotemporal Integration of Phonological and Semantic Processes: An Electroencephalographic Study. *The Journal of Neuroscience,* 32**,** 4297-4306.

ZATORRE, R. J., BOUFFARD, M. & BELIN, P. 2004. Sensitivity to auditory object features in human temporal neocortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience,* 24**,** 3637-42.

ZIEGLER, J. C., BENRAISS, A. & BESSON, M. 1999. From print to meaning: an electrophysiological investigation of the role of phonology in accessing word meaning. *Psychophysiology,* 36**,** 775-85.